Measuring precise radial velocities and cross-correlation function line-profile variations using a Skew Normal distribution

U. Simola X. Dumusque Jessi Cisewski-Kehe

May 24, 2018

Abstract

Stellar activity is the main limitation to the detection of small-mass exoplanets using the radial-velocity (RV) technique. Stellar activity can be probed by measuring variations in the shape of the cross-correlation function (CCF) as a function of time. Those variations are measured using the different moments of the CCF. Therefore measuring with the best precision those moments is crucial to de-correlate exoplanet signals from spurious RV signals originating from stellar activity.

We propose here to measure those moments using a Skew Normal (SN) distribution, which compared to the Normal distribution generally used, automatically includes an extra parameter to model the natural asymmetry of the CCF induced by convective blueshift.

We analyze 5 stars with different activity levels and whose CCF's have different signal-to-noise ratio (SNR) levels. In each case, we compare the results obtained by fitting to the CCF respectively a Normal and a SN. We also estimate rigorous errors for the different moments of the CCF using a bootstrap analysis.

The correlation between the RV's and the asymmetry of the CCF and the correlation between the RV's and the width of the CCF are always stronger when using the parameters derived from the SN in the case of real observations. Therefore the CCF asymmetry and the CCF width derived using a SN are more sensitive to stellar activity, which allows to probe with a better precision stellar rotational periods, in addition to characterize more precisely stellar activity signals. The precision on the estimated set of RV's, derived using the median of the fitted SN distributions are on average 10% smaller than the uncertainties calculated on the mean of the Normal. In addition, the uncertainties related to the asymmetry parameter derived from the SN are on average 15% smaller than the ones calculated on the common Bisector Inverse Slope Span (BIS SPAN). We strongly encourage the use of the SN distribution rather than the Normal distribution, because this allows us to retrieve in one single fit the different moments of the CCF, because the derived moments better catch stellar activity signals and finally because the standard errors on the RV's and the the asymmetry parameter are smaller than the one estimated with the Normal fit.

1 Introduction

When working with radial velocities (RV's), the main limitation to the detection of small-mass exoplanets is not anymore the precision of the instruments used, but the different noises induced by the stars we are observing (Dumusque et al. 2017). Indeed, stellar oscillations, granulation phenomena and stellar activity all induce RV signals (Saar and Donahue 1997; Queloz et al. 2001; Desort et al. 2007; Dumusque et al. 2011; Dumusque 2016) that are beyond the meter-per-second precision reached by the best high-resolution spectrographs. It is therefore mandatory to understand better stellar signals and to find ways to correct for them, if in the near future we want to detect or confirm an Earth-twin planet using the RV technique. This is even more true now that instrument like ESPRESSO (Pepe et al. 2014) and EXPRESS (Fischer et al. 2016) should have the stability to detect such signals. However, if solutions are not found to mitigate the impact of stellar activity, the detection or confirmation of potential Earth-twins will be extremely challenging.

Among the different stellar signals we are aware of, the one that is the most difficult to characterize and to correct for is the signal induced by stellar activity. Stellar activity is responsible for creating magnetic regions on the surface of stars, and those regions change locally the temperature and the convection, inducing spurious RV's variations (Meunier et al. 2010; Dumusque et al. 2014). In theory, it should be easy to differentiate between the pure Doppler-shift induced by a planet, that will shift the entire stellar spectrum and stellar activity, that modifies the shape of spectral lines and by doing so create a spurious shift of the stellar spectrum (Saar and Donahue 1997; Hatzes 2002; Kurster et al. 2003; Lindegren and Dravins 2003; Desort et al. 2007; Lagrange et al. 2010; Meunier et al. 2010; Dumusque et al. 2014). However, on quiet GKM dwarfs, the main target for precise RV's measurements, stellar activity induce signals of a few $\mathrm{m\,s^{-1}}$. This corresponds physically to variations smaller than 1/100th of a pixel on the detector.

To be able to measure such tiny variations, we average the information of all the lines in the spectrum by cross correlating the stellar spectrum with a synthetic (Baranne et al. 1996; Pepe et al. 2002) or an observed stellar template (Anglada-Escudé and Butler 2012). The result of this operation gives us the cross-correlation function (CCF). To measure the Doppler-shift between different spectra and therefore to retrieve the RV's of a star as a function of time, we calculate the variations of the CCF barycenter. The barycenter is generally estimated by fitting a Normal distribution to the CCF and taking its mean. Variations in line shape between different spectra, which indicate the presence of signals induced by stellar activity, is measured by analysing the different moments of the CCF. Usually, the width of the CCF is estimated using the FWHM of the fitted Normal distribution, and its asymmetry using the the bisector inverse slope span (BIS SPAN, Queloz et al. 2001).

If a RV signal is induced by activity, generally a strong correlation will be observed between the RV and chromospheric activity indicators like $\log(\mathrm{R}'_{HK})$ or H-α (Boisse et al. 2009; Dumusque et al. 2012; Robertson et al. 2014), but also between the RV and the FWHM of the CCF or its BIS SPAN (Queloz et al. 2001; Boisse et al. 2009; Queloz et al. 2009; Dumusque 2016). It is therefore common now, that when fitting for a planetray signal, in addition to a Keplerian, the model includes in addition linear dependancies with the $\log(\mathrm{R}'_{HK})$, the FWHM and the BIS SPAN (Dumusque et al.

2017; Feng et al. 2017). It is also common to add a Gaussian process to the model to account for the correlated noise induced by stellar activity. The hyperparameters of the Gaussian process are generally trained on the different activity indicators (Haywood et al. 2014; Rajpaul et al. 2015). It is therefore essential for mitigating stellar activity to obtain activity indicators that are the most correlated to the RV's but also for which we can obtain the best precision.

Several works have derived indicators that are more sensitive to line asymmetry than the BIS SPAN. In Boisse et al. (2011), the authors develop V_{span}, a new indicator to derive the CCF asymmetry that is more sensitive than the BIS SPAN at low signal-to-noise ratio (SNR). Figueira et al. (2013) studied the use of two new indicators, bi-Gauss and V_{asy}. The authors were able to show that when using bi-Gauss, the amplitude in asymmetry is 30% larger, therefore allowing the detection of smaller-amplitude correlations with RV's variations. They also demonstrated that V_{asy} seems to be the best indicator of line asymmetry at high SNR, as its correlation with RV is more significant than any other asymmetry indictors.

In all the methods described above, the RV and the FWHM are derived using a Normal distribution fitted to the CCF, and the asymmetry is estimated using another approach.

In this paper we propose to use a Skew Normal (SN) distribution to derive at the same time the RV, the FWHM and the asymmetry of the CCF, as this function includes a skewness parameter (Azzalini 1985). In addition, we know that for solar-type stars and cooler dwarfs, the bisector of the CCF as a "C"-shape due to convective blueshift (Dravins et al. 1981; Gray 2009). Therefore, fitting the CCF using a model that naturally includes an asymmetry, like the SN distribution, should give in principle more precise results.

The paper is organized as follow. In Sec. 2 we introduce the SN distribution and describe its applicability for modeling the CCF, showing that the SN distribution is a better representation of observed CCF than a Normal distribution, and study how the SN parameters relate to the RV, FWHM and BIS SPAN of the CCF. In Sec. 3 we present a simple model to correct from stellar activity which extends the ones previously proposed (Dumusque et al. 2017; Feng et al. 2017). In Sec. 4 we study the SN fit to the CCF using simulations coming from SOAP 2.0. In Sec. 5, we compare on real observations the sensitivity of the SN parameters to stellar activity with respect to other existing indicators. In Sec. 6 we derive error bars for the different CCF moments and finally we discuss our results and conclude in Sec. 7 and Sec. 8.

2 The Skew Normal distribution

The Skew Normal (SN) distribution is a class of probability distributions which includes the Normal distribution as a special case (Azzalini 1985). The SN distribution has, in addition to a location and a scale parameter analogous to the Normal distribution's mean and standard deviation, a third parameter which describes the asymmetry, or the skewness, of the distribution. Considering a random variable $Y \in \mathbb{R}$ (where $\mathbb{R}$ is the real line) which follows a SN distribution with location parameter $\xi \in \mathbb{R}$, scale parameter $\omega \in \mathbb{R}^+$ (i.e., the positive real line), and skewness parameter

$\alpha \in \mathbb{R}$, its density at some value $Y = y$ can be written as

$$SN(y; \xi, \omega, \alpha) = \frac{2}{\omega}\phi\left(\frac{y-\xi}{\omega}\right)\Phi\left(\frac{\alpha(y-\xi)}{\omega}\right), \tag{1}$$

where ϕ and Φ are respectively the density function and the distribution function of a *standard* Normal distribution[1] and $\alpha \in \mathbb{R}$ is the skewness parameter which quantifies the asymmetry of the SN. We then write $Y \sim SN(\xi, \omega^2, \alpha)$ to mean that the random variable Y follows the noted SN distribution. Examples of SN densities under different skewness parameter values and the same location and scale parameters ($\xi = 0$ and $\omega = 1$) are displayed in Fig. 1. A usual Normal distribution is the special case of the SN distribution when the skewness parameter, α, is equal to 0.[2] For reasons related to the interpretation of the parameters in Eq. (1) and computational issues

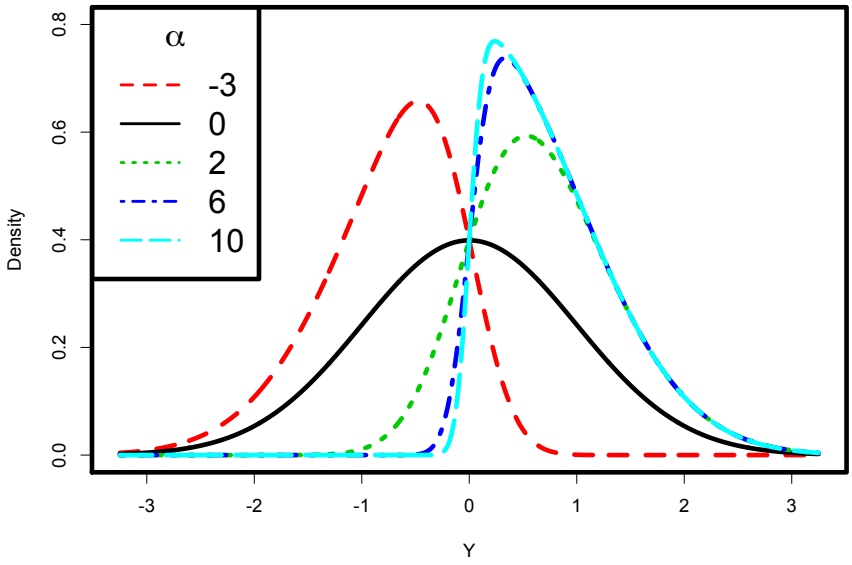

Figure 1: Density function of a random variable $Y \sim SN(\xi, \omega^2, \alpha)$ with location parameter $\xi = 0$, scale parameter $\omega = 1$ and different values of the skewness parameter α indicated by different colors and line types. Note that the solid black line has an $\alpha = 0$, making it a Normal distribution.

with estimating α near 0, a different parametrization is used, which is referred to as the *centered parametrization* (CP). We will be using the CP in this work, which includes a mean parameter μ, a variance parameter σ^2 and a skewness parameter γ. In order to define the CP, we need to express

[1] A standard Normal distribution is a Normal distribution with a mean of 0 and a standard deviation of 1.

[2] This can be seen from Eq. (1). If $\alpha = 0$ then $\Phi\left(\frac{\alpha(y-\xi)}{\omega}\right) = \Phi(0)$; this is the the probability a standard Normal random variable is less than or equal to 0, which is 0.5. The 0.5 cancels with the 2 in the denominator and what remains is the usual Normal density, $\frac{1}{\omega}\phi\left(\frac{y-\xi}{\omega}\right)$

the CP parameters (μ, σ^2, γ) as a function of the one used in the Equation (1) with (ξ, ω^2, α) by

$$\mu = \xi + \omega\beta, \quad \sigma^2 = \omega^2(1-\beta^2), \quad \gamma = \frac{1}{2}(4-\pi)\beta^3\left(1-\beta^2\right)^{-3/2}, \tag{2}$$

where $\beta = \sqrt{\frac{2}{\pi}}\left(\frac{\alpha}{\sqrt{1+\alpha^2}}\right)$.

By using Eq. (2), the new set of parameters (μ, σ^2, γ) provides a more clear interpretation of the behavior of the SN distribution. For the α values used in Fig. 1, the corresponding values of (μ, σ^2, γ) are displayed in Table 1. In particular, μ and σ^2 are the actual mean and variance of the distribution (rather than simply a location and scale parameter) and γ becomes an index for evaluating the skewness of the SN.

Beyond the mean of the SN, we introduce a second location parameter that will be largely used in the analyses: the median. The median of the SN, and in general the median of an absolute continuous random variable, is defined as that value m such that

$$\int_{-\infty}^{m} SN(y; \xi, \omega, \alpha) = \frac{1}{2}, \tag{3}$$

where $SN(y; \xi, \omega, \alpha)$ follows Eq. (1).[3]

α	μ	σ^2	γ
-3	-0.757	0.427	-0.667
0	0.000	1.000	0.000
2	0.714	0.491	0.454
6	0.787	0.381	0.891
10	0.794	0.370	0.956

Table 1: CP values, (μ, σ^2, γ), corresponding to the α values from Fig. 1 (with location parameter $\xi = 0$ and scale parameter $\omega = 1$) using Eq. (2). Values are rounded to three decimal places.

Further details about the parametrization from Eq. (1) (called *Direct Parametrization* or DP), the CP, and general statistical properties of the SN are treated in rigorous mathematical and statistical viewpoints in the book by Azzalini and Capitanio (2014).

2.1 Fitting the Skew Normal distribution to the CCF

The SN density shape is used to model the CCF. In particular, using least-squares algorithm, we fit the following function:

$$f_{CCF}(y_i) = \mathrm{C} - \mathrm{A} \times SN(y_i; \mu, \sigma^2, \gamma), \quad i = 1, \ldots, n \tag{4}$$

[3]We recall that when using a symmetric distribution such as the Normal distribution, the mean and the median are equivalent.

where C is an unknown offset fitting the continuum of the CCF, A is an unknown amplitude parameter known with the term "contrast" and $y_1, \ldots, y_n$ are the set of RV's considered for the CCF. Note that the CCF is expressed in flux as a function of the lag of the cross-correlation template, expressed in RV.

Since the CCF presents a natural asymmetry due the convective blueshift, the SN distribution should in principle better catch this aspect, together with other changes in asymmetry, respect the Normal fit. To initially check this intuition, we compared the CCF residuals after fitting a Normal and a SN distribution for 2 stars. The first star is Alpha Centauri b, whose CCF's have high signal-to-noise ratio (SNR). The second star is Corot-7, whose CCF's have low SNR. Fig. 2 shows that the SN seems to be a slightly better model to explain the shape of the CCF, in particular as the SNR decreases.

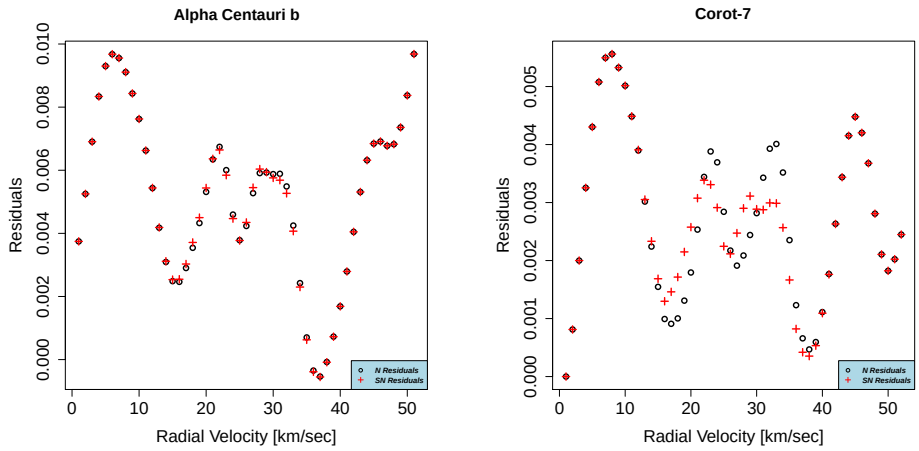

Figure 2: Comparison between the Normal (black circles) and the SN (red crosses) residuals using CCF's from the star Alpha Centauri b (left) and Corot-7 (right). When looking at the residuals corresponding to the tails of the CCF, the results derived by the two fits are comparable. However, when focusing on the center of the CCF, the SN fit leads to slightly better results for both the stars. Moreover, as the SNR decreases, the SN distribution shows smaller residuals respect the Normal ones.

In the following of the paper, we define RV as the mean of the Normal distribution. Concerning the fit of the CCF using the SN, we present at first 2 indicators that define the RV of the star: the mean of the SN, defined as SN mean RV and the median of the SN, defined as SN median RV (i.e. realling Eq.(3), m = SN median RV). We will discuss advantages and limits for both these choices in Sec. 5 and Sec. 6. For the width of the CCF, we use the FWHM of the Normal, which is $2\sqrt{(2\ln 2)}\sigma$. The width of the SN, SN FWHM, is defined in the same way[4]. Being

[4]Note that SN FWHM does not correspond to the width of the SN distribution at half maximum like in the Normal case.

a Normal distribution symmetric, there is not such a parameter that evaluates the asymmetry of the distribution, so the BIS SPAN is used. The BIS SPAN will be compared to the asymmetric parameter γ of the SN, known also as SN GAMMA. To test the strength of the correlation between the estimated RV's and the different indicators used to evaluate stellar activity, we calculated the Pearson correlation coefficient, which in its general form is defined as:

$$R(x, y) = \frac{cov(x, y)}{\sigma(x), \sigma(y)}, \tag{5}$$

where x and y are two quantitative variables, $cov(x, y)$ indicates the covariance between x and y, and $\sigma(x)$ and $\sigma(y)$ represent their standard deviations. A $p-$value for the statistical test having null hypothesis $H_0 : R = 0$ is provided, along with a 95% confidence interval for R when needed.

3 Radial Velocity correction function for stellar activity

Exoplanets will only produce a variation in RV's induced by a pure Doppler-shift of stellar spectra. Stellar activity and in particular the presence of active regions on the photosphere of the star, on the contrary, does not produce a blueshift or redshift of the spectra, but creates spurious RV's signal by modifying the shape of the spectral lines. To track these changes in the shape of the line profile, the general approach consists in using the FWHM, the BIS SPAN or the indicators introduced by (Figueira et al. 2013), which provide an information on the average width and asymmetry of the CCF. A strong correlation between the estimated set of RV's and one or more of these parameters provides an indication that the RV's are affected by stellar activity signals.

Because of the spurious variations in RV's caused by active regions, a crucial step of the entire analysis is to correct the initially estimated RV's in order to retrieve a set of RV's containing ideally only the possibly pure Doppler-shift coming from the exoplanets. In order to reach this goal, we propose to consider a linear combination of the RV's having as covariates the amplitude parameter A, the BIS SPAN (or γ), the FWHM (or SN FWHM) and the interaction between the BIS SPAN and the FWHM (or γ and SN FWHM and the interaction between γ and SN FWHM in the SN case). We propose the following function to correct the RV's from stellar activity:

$$RV_{\text{stellar activity}} = \beta_0 + \beta_1 A + \beta_2 \gamma + \beta_3 \text{SN FWHM} + \beta_4 (\gamma \text{SN FWHM}) + \epsilon, \tag{6}$$

where β_0 is the intercept and ϵ is the vector of the errors with mean equal to 0 and covariance matrix equal to $\sigma^2 I$ (I defined as the identity matrix).

While the parameters BIS SPAN and FWHM have already been introduced and used in (Dumusque et al. 2017; Feng et al. 2017), by using Eq. (6) we propose also to include the contrast parameter A and a variable that quantifies the interaction between BIS SPAN (or γ) and FWHM (or SN WHM). The inclusion of the contrast parameter A is justified by the expected behavior of the CCF in presence of active regions. Fig. 3 shows that the presence for instance of a spot produces also changes in the amplitude of the CCF and not only on its asymmetry or width.

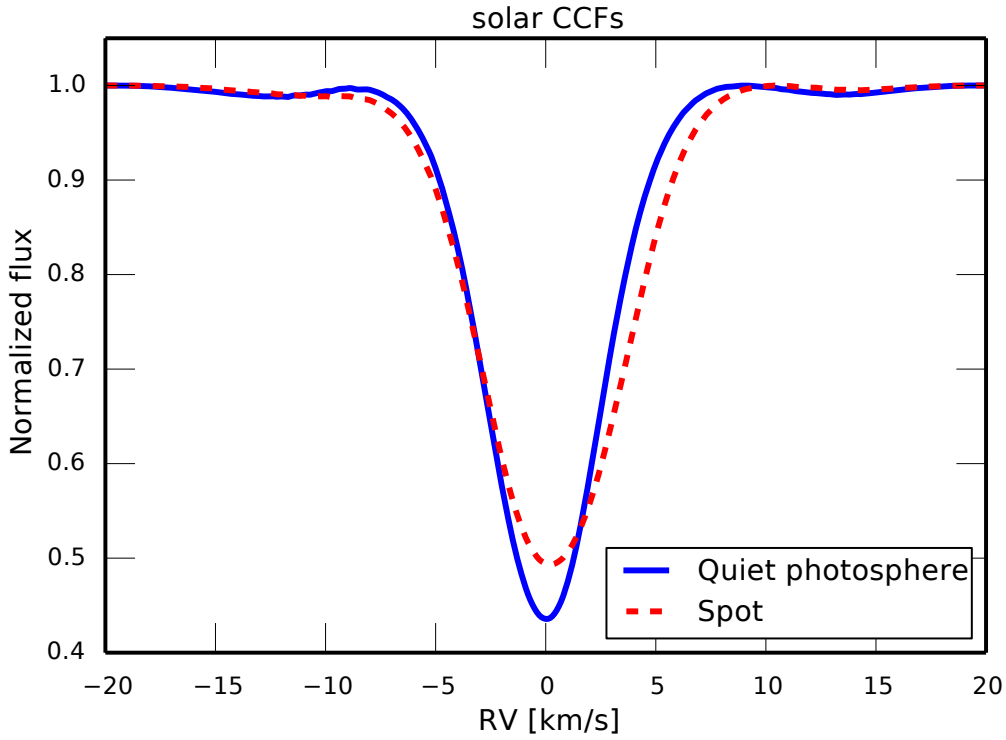

Figure 3: The presence of a spot on the photosphere of a star produces changes on the amplitude of the CCF (also known as contrast) and not only on its asymmetry or width. This very same plot has been originally presented by (Dumusque et al. 2014) when introducing the Spot Oscillation And Planet (SOAP) 2.0 simulator.

The reasons for including in Eq. (6) a variable that quantifies the interaction between BIS SPAN (or γ) and FWHM (or SN FWHM) will be better justified through the results of the examples presented in Sec.4. We point out however that while the association between in this case BIS SPAN (or γ) and FWHM (or SN FWHM) means that the values of one variable relate to the values of the other (since in this case we have two quantitative variables we talk about correlation), with the term interaction we mean that the effect that one variable has on the RV's is not constant. In particular the effect differs at different values of the other values.

In order to show the goodness of this correction, a statistical test on the parameters $\beta_0, \beta_1, \beta_2$, β_3 and β_4 is presented, where the null hypothesis is $H_0 : \beta_i = 0$, for $i = 0, \ldots, 4$. The level for not rejecting the null hypothesis is fixed equal to 0.05. The coefficient of multiple correlation R^2 is introduced in order to explain how well this linear combination addresses the variability of the RV's of the star as caused by stellar activity.

As final remark we recall that, when working with a linear regression, there are several ways to select the variables to include in the model. While usually the stepwise technique takes place (Efroymson 1960; Hocking 1976), the proposed function defined in Eq. (6) that corrects from stellar activity is the result of statistical and astronomical considerations. In particular we checked that the correlations between the proposed parameters were not approaching one: if it were the case, the matrix needed to calculate the estimates would have been singular, hence non invertible. This problem is known in statistic with the term multicollinearity. A detailed discussion of the topic can be found in the book by (Belsley 1991). In the literature there are examples of stars presenting a high correlation between the contrast parameter A and the FWHM (Dumusque et al. 2017; Feng et al. 2017). In the analyses of the presented work, we did not find any correlation between the contrast parameter A and the FWHM (or between the contrast parameter A and the SN FWHM) exceeding 0.66 for all the analyzed stars and therefore the problem of multicollinearity is avoided. For the interpretation of the coefficients, we centered on the mean all the variables considered in the linear regression defined in Eq. (6). Last, when proposing interactions between covariates, we found that only the one between the width and the asymmetry of the CCF is statistically helpful to explain spurious variations in RV's caused by stellar activity. Consequently, in the present work, the interaction between other variables are not used as covariates.

4 Simulation Study

In order to evaluate the performance of the proposed SN approach for modeling the CCF and the goodness of the proposed correction from stellar activity by using Eq. (6), we begin by considering a simulation study using spectra generated from the Spot Oscillation And Planet (SOAP) 2.0 code (Dumusque et al. 2014).

[[**Umberto:** Add more information on SOAP, the way we retrieve the CCFs with SOAP, the characteristics of the CCF (i.e. how many points, ...) and the reasons why we need CCFs simulated with SOAP.]] [[**Umberto:** Specify characteristic of the star.]] [[**Xavier:**]]

4.1 Faculae

[[**Umberto:** Specify characteristic of the faculae.]][[**Xavier:**]]

Fig. 4 shows the results obtained when a faculae is present on the photosphere of the star. It is possible to note that, although there is not a planet, the presence of the faculae leads to spurious variations in RV's for all the proposed indicators. SN mean RV seems to have the smallest spurious variations caused by the faculae.

Since in this case we know that the variations in RV are only caused by the faculae, the evaluation of the correlations between the set of RV's and the asymmetry parameters can provide relevant information to understand the cause of the variations observed in RV's. Fig. 5 shows the correlations obtained using as RV respectively SN mean RV, SN median RV and RV. The γ parameter shows a correlation with SN median RV equal to -0.944, which is stronger than the correlations between the other asymmetry indicators and their estimated RV's. When focusing on the width of the CCF, The correlation between SN FWHM and SN mean RV is the strongest one ($R^2 = 0.655$).

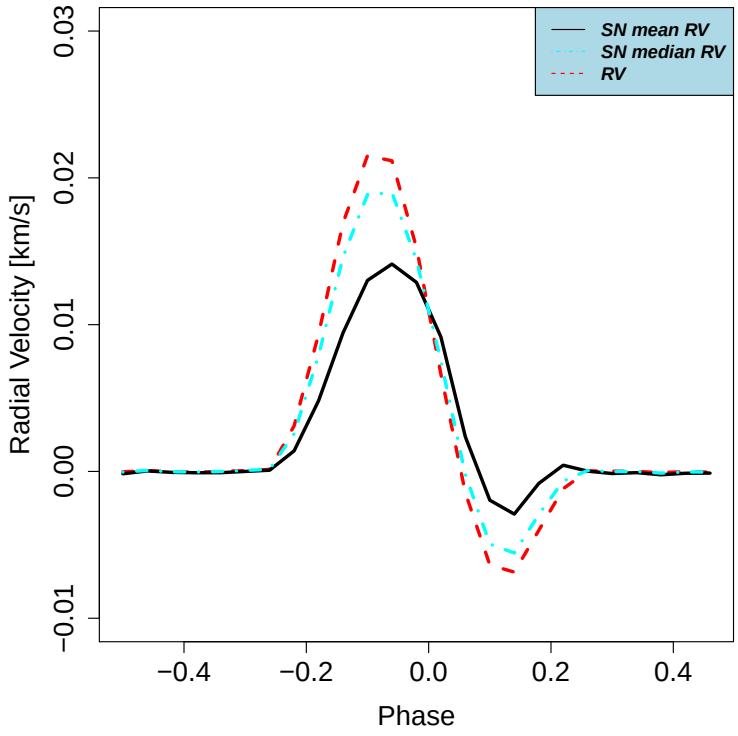

Figure 4: RV's changes as function of the orbital phase in the case in which a faculae is present on the photosphere of the star. SN mean RV seems to have the smallest spurious variations caused by the faculae.

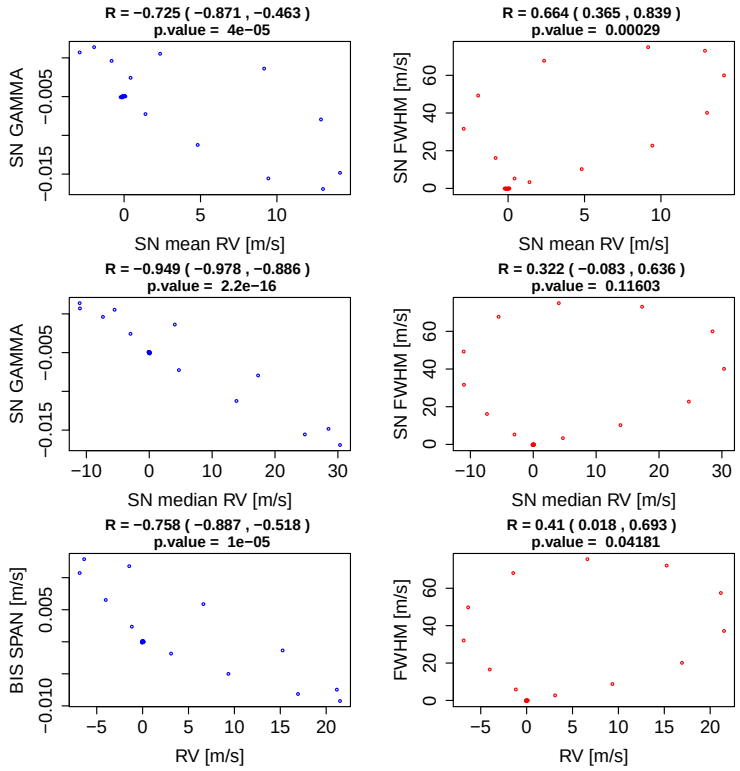

Figure 5: Evaluation of the correlation between the RVs and the asymmetry parameters when a faculae is present on the photosphere of the star. In this case both the shape and the width of the CCF changes as the faculae moves, producing statistically significative correlations between the RVs and respectively the asymmetry parameter and the width parameter.

Since all the variations in RV's displayed in Fig. 4 are caused by the presence of a faculae on the photosphere of the star, we corrected the originally estimated set of RV's by using Eq. (6). The results of the correction are displayed in Fig. 6 and the statistical tests on the parameters involved in Eq. (6) are summarized in Table 2. It is straightforward noticing that the proposed correction almost completely addresses the issue, providing a new set of RV's where there are not anymore spurious variations caused by the presence of the faculae. In particular the R^2 is extremely high regardless the fit that has been use (Normal or SN). We note also that the parameter β_4, associated with the interaction between γ and SN FWHM (or BIS SPAN and FWHM) is not statistically significant to explain spurious variations caused by a faculae.

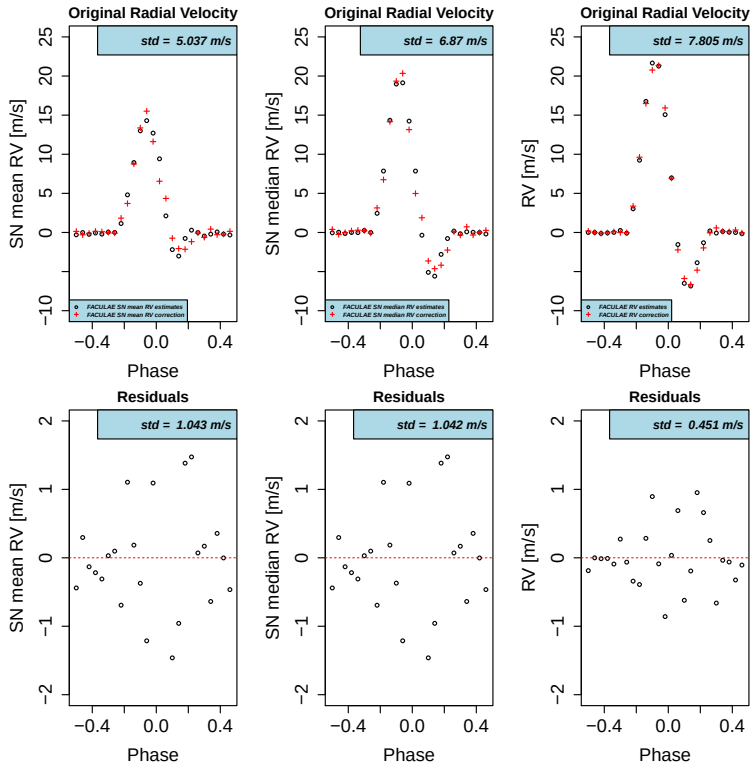

Figure 6: Set of spurious variations in RV's caused by a faculae using a Normal and a SN fit before and once corrected from stellar activity. The correction is done using Eq. (6) and the estimated parameters are presented in Table 2. Once corrected for stellar activity the residuals do not show a systematic component.

4.2 Spot

[[**Umberto:** Specify characteristic of the spot.]][[**Xavier:**]]

Parameter	RV	SN mean RV	SN median RV
β_0	0.0016	0.26	0.60
β_1	$3.75e-05$	0.0068	0.0069
β_2	$7.79e-16$	$1.97e-05$	$3.69e-08$
β_3	0.25	0.0011	0.0011
β_4	0.53	0.21	0.21
R^2	0.9967	0.9571	0.977

Table 2: Evaluation of the linear combination used for correcting the RV's from spurious variations caused by a faculae, according to Eq. (6). Beyond the contrast parameter A and the asymmetry parameter of the CCF which are useful for both the analyses, the width of the CCF is statistically useful to explain spurious variations in RV's caused by a faculae only for the SN analyses. The term of interaction between the asymmetry and the width of the CCF is not helpful to explain part of the variability in the set of RV's. The estimated R^2 show that the proposed correction for stellar activity explains all the spurious variability in RV's.

Fig. 7 shows the results obtained when a spot is present on the photosphere of the star. Similarly to the previous case, the presence of the spot leads to spurious variations in RV's for all the proposed indicators, but SN mean RV seems to have the smallest ones. We note however that in this case all the three indicators are similarly effected by the spurious changes in RV's caused by the spot.

Fig. 8 shows the correlations between the asymmetry parameters and respectively SN mean RV, SN median RV and RV. The correlation between γ and SN median RV is equal to -0.89 while the correlation between γ and SN median RV is equal to -0.82. Both these correlations are stronger than the correlation between the BIS SPAN and RV ($R^2 = -0.77$). The correlations between the width of the CCF and the corresponding RV's are slightly stronger when fitting a SN rather than a Normal distribution, as shown in the left series of plots in Fig. 8. The results show anyway a correlation close to 0 when focusing on the width of the CCF, suggesting that the presence of a spot on the photosphere of star leads to changes in the shape of the CCF rather than its width.

As before, we corrected the originally estimated set of RV's by using Eq. (6). The results of the correction are displayed in Fig. 9 and the statistical tests on the parameters involved in Eq. (6) are summarized in Table 3. Also in this case the proposed correction almost completely addresses the issue, providing a new set of RV's where there are not anymore spurious variations caused by the presence of the spot. The R^2 is extremely high regardless the fit that has been use (Normal or SN). The main differences respect to the previous case are that the parameter β_3, associated with γ, and the parameter β_4, associated with the interaction between γ and SN FWHM (or BIS SPAN and FWHM), are statistically significant to explain spurious variations caused by a spot. This information could be useful in the future for disentangle spots from facuale.

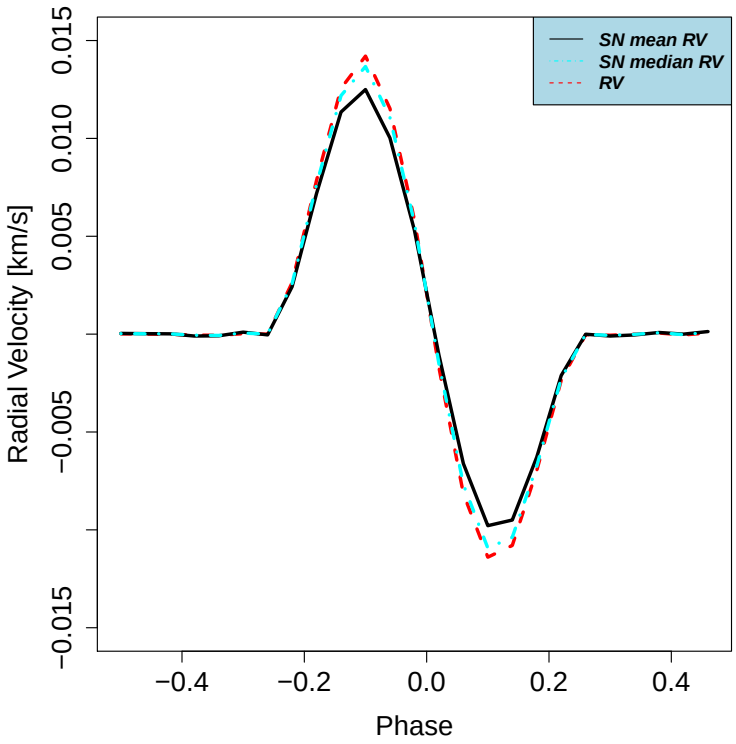

Figure 7: RV's changes as function of the orbital phase in the case in which a spot is present on the photosphere of the star. SN mean RV seems to have the smallest spurious variations caused by the faculae.

Parameter	RV	SN mean RV	SN median RV
β_0	0.47	0.84	0.81
β_1	0.00035	0.064	0.064
β_2	$2e-16$	$2e-16$	$2e-16$
β_3	0.53	0.61	0.61
β_4	$1.11e-07$	$2.85e-08$	$2.87e-08$
R^2	0.9897	0.9914	0.9929

Table 3: Evaluation of the linear combination used for correcting the RV's from spurious variations caused by a spot, according to Eq. (6). All the covariates are statistically useful to explain the variability in RV's causued by a spot except the intercept and the width of the CCF. Concerning the SN analysis, the contrast parameter A is not statistically significant at level 5%. The estimated R^2 show that the proposed correction for stellar activity explains all the spurious variability in RV's.

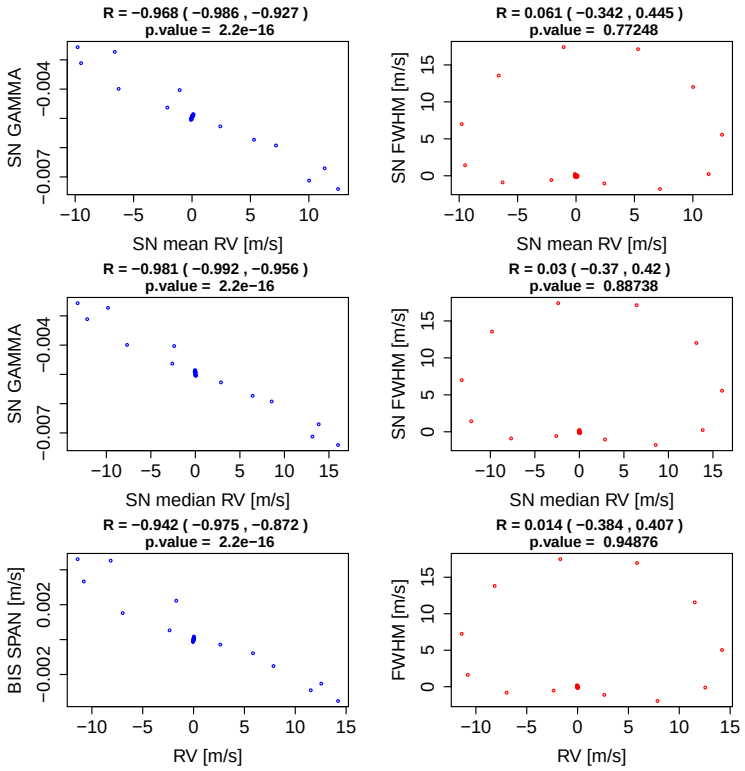

Figure 8: Evaluation of the correlation between the RV's and the asymmetry parameters when a spot is present on the photosphere of the star. In this case only the shape of the CCF changes as the spot moves, producing statistically significative correlations only between the RV's and the asymmetry parameter.

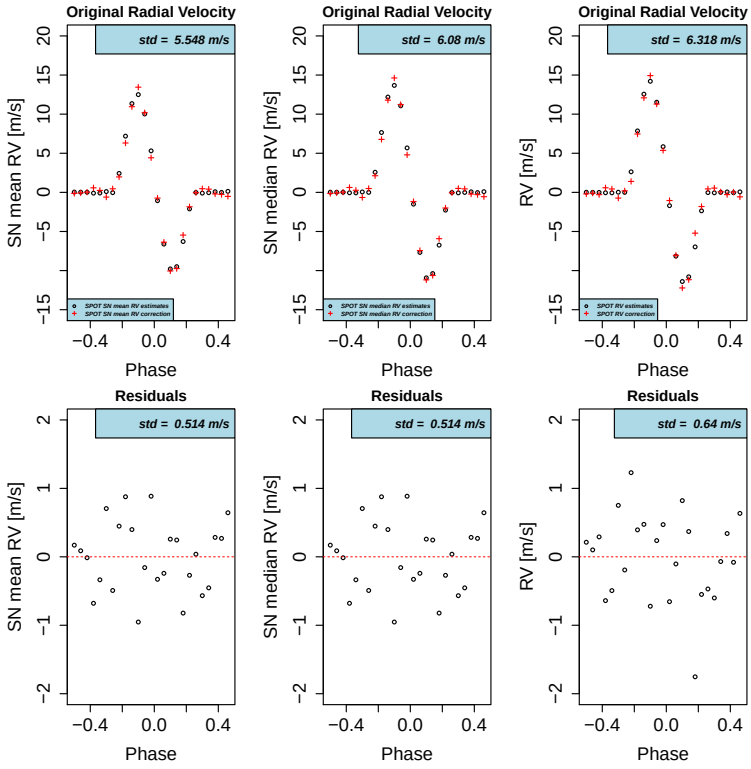

Figure 9: Set of spurious variations in RV's caused by a spot using a Normal and a SN fit before and once corrected from stellar activity. The correction is done using Eq. (6) and the estimated parameters are presented in Table 2. Once corrected for stellar activity the residuals do not show a systematic component.

4.3 Spot and planet

The last presented simulation example consists in having not only spurious variations in RV's caused by the presence of a spot on the photosphere of the star but also a pure Doppler-shift because of an artificially injected planet. The purpose of this example is to check if we are able to disentangle the two different sources of variations in the set of RV's.

Fig. 10 shows the results obtained when a spot is present on the photosphere of the star and a planet is injected. The planet, having an amplitude of 10 $\mathrm{m\,s^{-1}}$, produces a pure Doppler-shift on the CCF, without further changing its shape. In this case N mean RV seems to have the largest variations caused by the combined action of spot and planet.

Fig. 11 shows the correlations between γ and respectively SN mean RV and SN median RV and the correlation between BIS SPAN and RV. In this case the correlations are weaker than the ones derived when only a spot is present on the photosphere of the star. Anyway the γ parameter shows a correlation with the median SN of -0.433, which is stronger than what the correlation between the other asymmetry indicators and their corresponding RVs. The correlations between the width of the CCF and the corresponding RV's are comparable and, similarly to the previous case, close to 0.

In order to correct the estimated set of RV's from spurious variations caused by the spot, we need to generalize Eq. (6), in order to consider also the possibly presence of an exoplanet. Hence, rather than just using Eq. (6) we add the following keplerian function, which represents the pure Doppler-shift of the exoplanet under the assumption of circular orbit:

$$RV_{\text{exoplanet}} = K \sin\left(\frac{2\pi}{P}(t - t_0)\right), \tag{7}$$

where the amplitude K, the orbital period P and the epoch at the periapsis t_0 are three unknown parameters that define the presence of the exoplanet. Overall, the variations in RV's of the star can be defined as

$$RV = RV_{\text{stellar activity}} + RV_{\text{exoplanet}}. \tag{8}$$

We note that the p–value associated with the amplitude parameter K is particularly relevant for rejecting or not rejecting the assumption about the presence of an orbiting companion. Moreover we note also that Eq. (7) is highly non linear, meaning that the estimation of all the parameters involved in Eq. (8) has to be done numerically. The statistical tests conducted on the parameters, whose results are summarized in Table 4, are significant. We note in particular that the p–value associate to the amplitude parameter K is $2e-16$. This result can be justify by the fact that the planet artificially injected has amplitude $K = 10\,\mathrm{m\,s^{-1}}$.

4.4 Conclusions on the simulation study

In this Sec.4 we presented a first implementation of the SN fit to CCF, using the simulation environment SOAP. Before moving to real cases, where the analyses on five stars are presented, we need to provide further considerations. First of all, looking ad the analyses conducted with SOAP, it seems

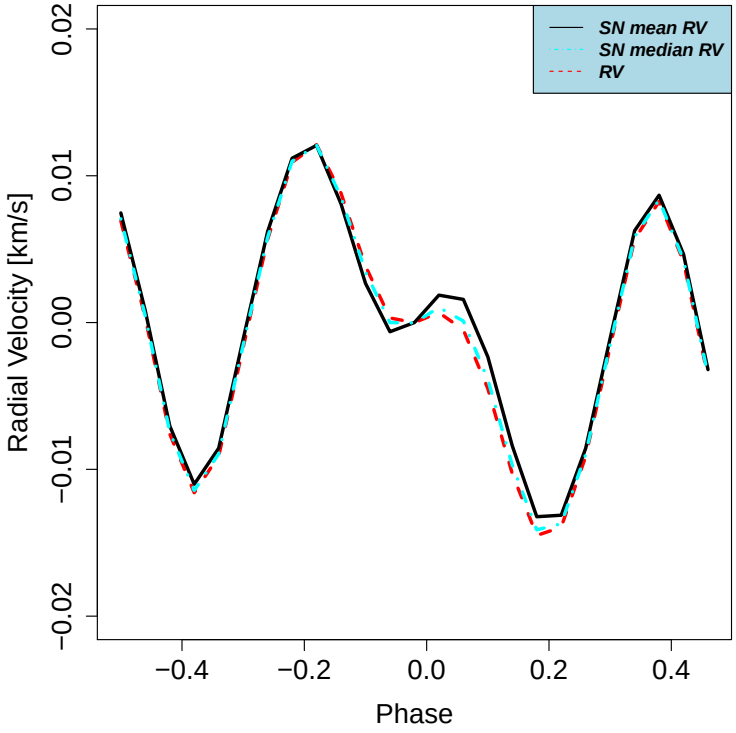

Figure 10: RV's changes as function of the orbital phase in the case in which a spot is present on the photosphere of the star and a planet is injected. N mean RV seems to have the largest variations caused by the combined action of spot and planet.

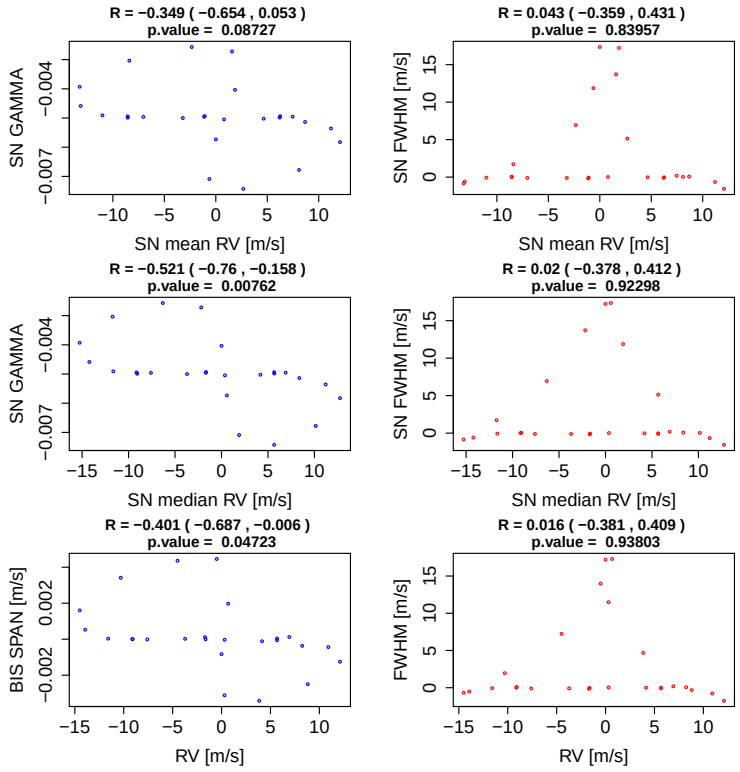

Figure 11: Evaluation of the correlation between the RV's and the asymmetry parameters when a spot is present on the photosphere of the star and a planet is injected. In this case only the shape of the CCF changes as the spot moves, producing statistically significative correlations only between the RVs and the asymmetry parameter. The correlations between the RVs and the width parameter of the CCF is weaker than the previous case that considers only the presence of a spot on the photosphere of the star.

Parameter	RV	SN mean RV	SN median RV
β_0	$7.09e-09$	$3.44e-07$	$4.96e-09$
β_1	0.00105	0.0052	0.0052
β_2	$2e-16$	$2.43e-16$	$2e-16$
β_3	0.79	0.22	0.22
β_4	$6.73e-07$	$7.78e-06$	$7.82e-06$
K	$2e-16$	$2e-16$	$2e-16$
P	$2e-16$	$2e-16$	$2e-16$
t_0	$2e-16$	$2e-16$	$2e-16$
Residuals	$0.61\,\mathrm{m\,s^{-1}}$	$0.60\,\mathrm{m\,s^{-1}}$	$0.60\,\mathrm{m\,s^{-1}}$

Table 4: Evaluation of the linear combination used for correcting the RV's, according to Eq. (8). All the parameters are statistically helpful to address spurious variations in RV's except the FWHM. Concerning the keplerian parameters, the amplitude K that provides relevance about the possibly presence of the exoplanet. Note that since non linear least squares are required, the residual standard error rather than the R^2 is displayed for each case.

that the largest correlation between an asymmetry parameter and a set of RV's happens to be when respectively γ and SN median RV are used. This is a bit surprising, since as the shape of the CCF changes, we expect SN median RV to be more robust than SN mean RV. [[**Umberto:** A possible justification of this ...]]. As second, when searching for stellar activity by deriving the correlation between the set of RV's and either an asymmetry parameter or the width of the CCF, the latter leads to weaker and hence less conclusive results if the active region is a spot. When stellar activity is dominated by faculae, both the shape and the width of the CCF changes as the faculae evolves on the photosphere of the star. Related to these last two considerations, we note that the interaction between the asymmetry and the width of the CCF is useful to explain part of the variability in the RV's if the active region is a spot but not when it is a faculae. The proposed function to correct for stellar activity addressed high level of spurious variations in RV's caused by active regions. In particular, respect to other common linear interpolation, we proposed to use as covariates also the amplitude parameter of the CCF and the interaction between γ and SN FWHM (or BIS SPAN and FWHM). As a consequence of using the interaction between the asymmetry and the width of the CCF, we note that the FWHM (or SN FWHM) becomes statistically not significant, while this is not the case if the interaction term is not involved in the linear regression. Finally, the correlations involving the common indicators (i.e. RV, FWHM and BIS SPAN) are systematically weaker than the correlations obtained by fitting the SN to the CCF, suggesting that this distribution could be helpful when searching for active regions. We recall moreover that all the quantities needed for conducting the analyses of the CCF are directly available by just fitting the SN.

5 Real data application

In this Section we present the analyses conducted on Alpha Centauri b, comparing the result of fitting a CCF using the SN distribution defined in Sec. 2.1 with the approach based on the Normal distribution. Other four stars have been analyzed with the proposed method and details can be found in the Appendix A. For all the stars that have been considered in the present work, we selected those CCF's having SNR larger than 10.

A comparison with the results obtained by the classic approach is done, where the RV's of the star are estimated by retrieving the mean of the Normal distribution used to fit the CCF, along with the BIS SPAN or the other asymmetric parameters defined in Figueira et al. (2013). The latter parameters are calculated separately from the Normal fit that leads to the set of RV's of the star.

5.1 Alpha Centauri B

A total of 1808 CCF's measured in 2010 have been analysed from the star Alpha Centauri B. Several measurement in 2010 are contaminated by light from Alpha Centauri A. To remove contaminated spectra and thus CCF's, we performed the same selection as presented in Dumusque et al. (2012). Moreover, as noted in Dumusque et al. (2012) and Thompson et al. (2017), this dataset presents a strong stellar activity signal.

We begin the analyses by evaluating the correlation between γ and the BIS SPAN. In Fig. 12, we see that the relationship between γ and the BIS SPAN is linear, with a slope equal to 0.72 and a strong Pearson correlation coefficient of 0.954. This comparison is useful because γ is an adimensional parameter taking information about the asymmetry of the SN while the BIS SPAN, beyond this, has got unit of measure of $\mathrm{km\,s^{-1}}$. In other words, by using Fig. 12, it is possible to provide a physical meaning to γ.

Fig. 13 shows the comparison between the RV's retrieved using the SN shape and the ones obtained with the Normal shape. It is possible to appreciate the presence of a strong stellar activity signal, as expected (Dumusque et al. 2012; Thompson et al. 2017). When using SN mean RV, it is possible to observe more variations than the ones measured by the Normal fitting. This happens because the mean of the SN is more sensitive to stellar activity. In fact, because the SN includes an asymmetry parameter, SN mean RV gets more shifted in the direction of the asymmetry induced by stellar activity. On the other hand, when using SN median RV, smaller variations in RV are caused by changes in the asymmetry of the CCF, because this second location parameter is a more robust indicator than the mean. The bottom plot of Fig. 13 captures this aspect. Both indicators can be used to capture and summarise the different information available in the CCF, as will be shown in the remainder of this work.

Similar to Figueira et al. (2013), we compare the correlation between the different activity indicators and the RV's of the star in Fig. 14. The correlation between γ and SN mean RV and the correlation between γ and SN median RV are much stronger than the correlations calculated between the other asymmetry parameters and their corresponding RV's. In particular the correlation between γ and SN mean RV is almost twice the correlation between the other asymmetry

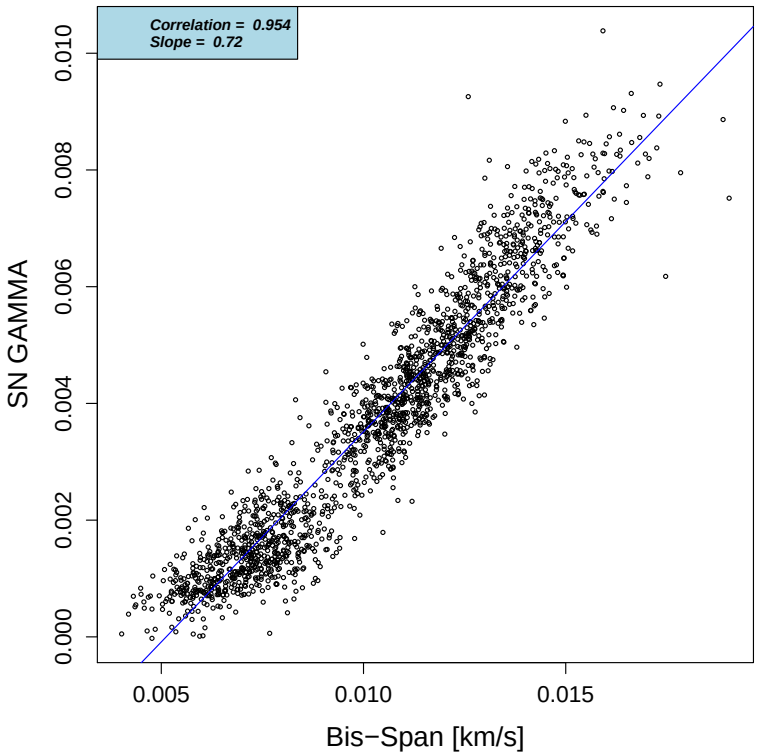

Figure 12: Correlation between γ and the BIS SPAN for Alpha Centauri B. Because γ is adimensional, retrieving the slope between γ and the BIS SPAN, which is expressed in $\mathrm{km\,s^{-1}}$, allows us to provide physical meaning to γ.

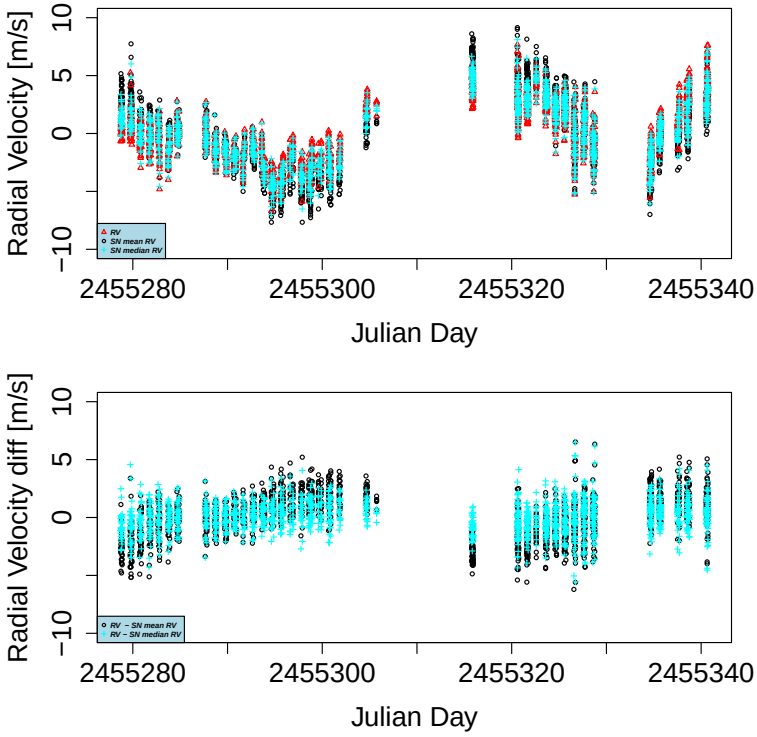

Figure 13: (top) RV's as function of Julian Day for Alpha Centauri b. The RV's are retrieved using the mean of the Normal (red triangles), SN mean RV (black circles), SN median RV (cyan crosses). (bottom) RV differences between Normal RV and SN mean RV (black circles) and between Normal RV and a SN median RV (cyan crosses).

Parameter	RV	SN mean RV	SN median RV
β_0	0.49	0.90	0.027
β_1	$2.22e-16$	$2.22e-16$	$2.22e-16$
β_2	0.33	$2.22e-16$	$1.23e-11$
β_3	$2.22e-16$	$2.22e-16$	$2.22e-16$
β_4	$2.22e-16$	$2.22e-16$	$2.22e-16$
R^2	0.57	0.78	0.66

Table 5: Evaluation of the linear combination used for correcting the RV's, according to Eq. (6). Concerning the Normal fit, all the parameters but the intercept and the BIS SPAN are useful in explaining variations in RV's of the star that can be caused by stellar activity. For the SN fit we note that the parameter related to γ is highly significant to address part of the spurious variations in RV's caused by stellar activity. The evaluation of the R^2 shows that the proposed linear combination better explains variations in RV's due to stellar activity coming from the SN analysis that uses SN mean RV.

parameters and their corresponding RV's.

Because the median is a more robust index than the mean, the correlation between γ and SN median RV is not as large as the correlation between γ and SN mean RV, but it is nonetheless 1.5 times larger than the correlation between the other common asymmetry parameters and their corresponding RV's. In other words, changes in the asymmetry of the CCF are better captured when using the SN mean RV. The correlation between FWHM and the RV's, either by using SN mean RV or SN median RV, is as well stronger when fitting a SN distribution rather than a Normal. All the correlations are statistically different from 0. Recalling the analyses presented in Sec. 4, we could infer that Alpha Centauri b is dominated by faculae, because the correlations between the RV's and the width of the CCF are strong (in particular the correlation between SN mean RV and SN FWHM is 0.817).

Using Eq. (6), we provide a new set of RV's corrected from stellar activity. The results are shown in Figure 15. We see that, once corrected for stellar activity, the residuals in the Normal and SN analysis are comparable. However, we note that when using SN mean RV, the correction is more important. In fact, the comparison of R^2 shows that the SN fit accounts for a higher percentage of variability caused by stellar activity (i.e. spurious variations in RV's). We note also that the BIS SPAN is not helpful to address part of the spurious variability caused by stellar activity, while the opposite conclusion has reached when evaluating the p-value related to γ.

Both the proposed indicators coming from the SN distribution have advantages and limits: SN mean RV better catches changes in the asymmetry of the CCF but the resulting set of RV's ends up being contaminated by those spurious shifts caused by stellar activity that have been shortly presented in Sec. 1. When using SN median RV, the final set of RV's is less affected by those spurious shifts caused by stellar activity, but at the same time this indicator is not able to catch as well as SN mean RV changes in the shape and in the width of the CCF. Once corrected from stellar activity using Eq. (6), the results are comparable. Anyway, both SN mean RV and SN median RV

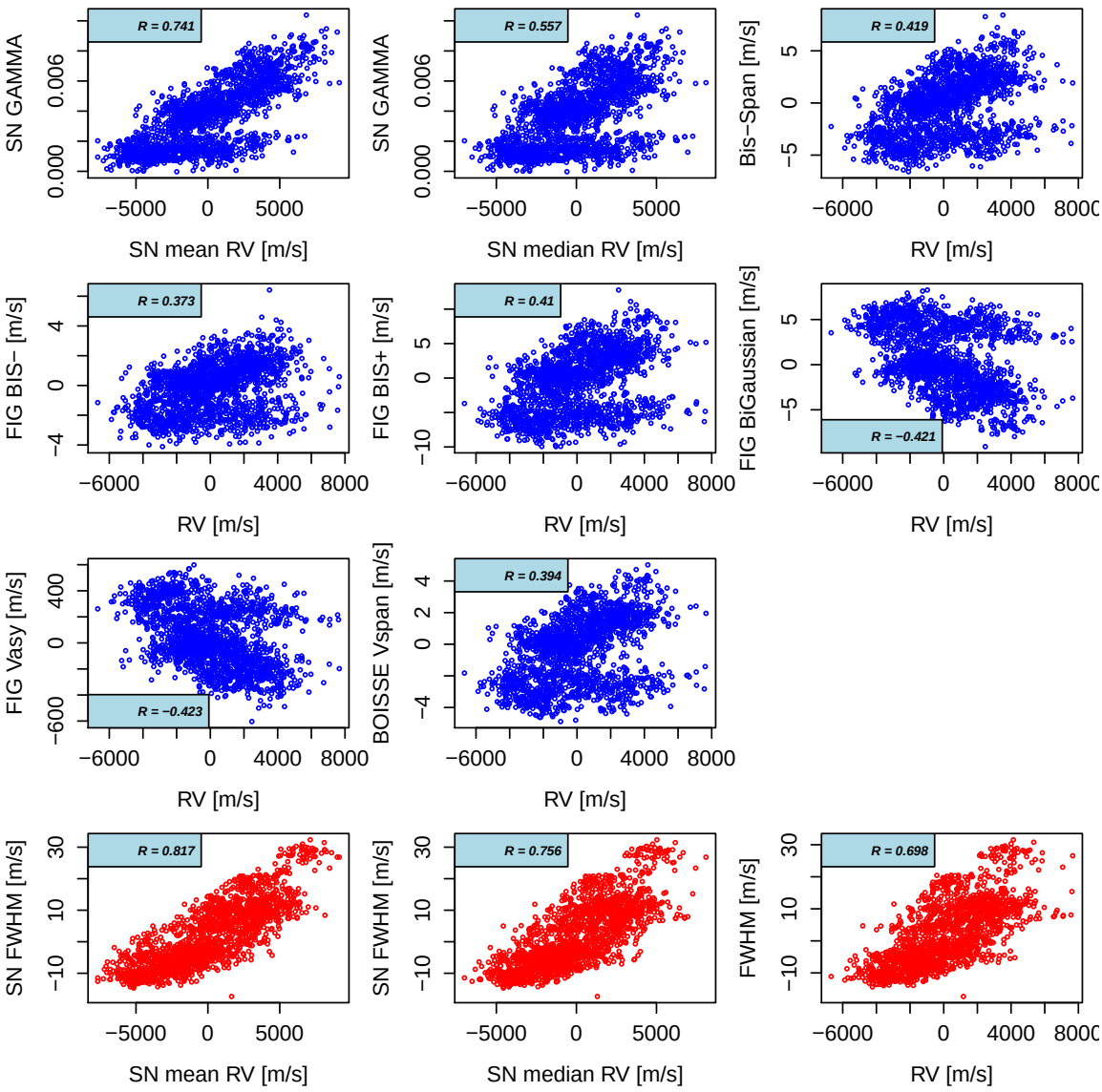

Figure 14: Correlation between the asymmetry parameters and the RVs for Alpha Centauri B. The last three plots show the correlation between the FWHM and the RVs for Alpha Centauri B using respectively the SN (SN mean RV and SN median RV) and the Normal fits. The correlations are always stronger when using parameters derived from the SN fit than the Normal one. The p–values associated with each R is statistically different from 0.

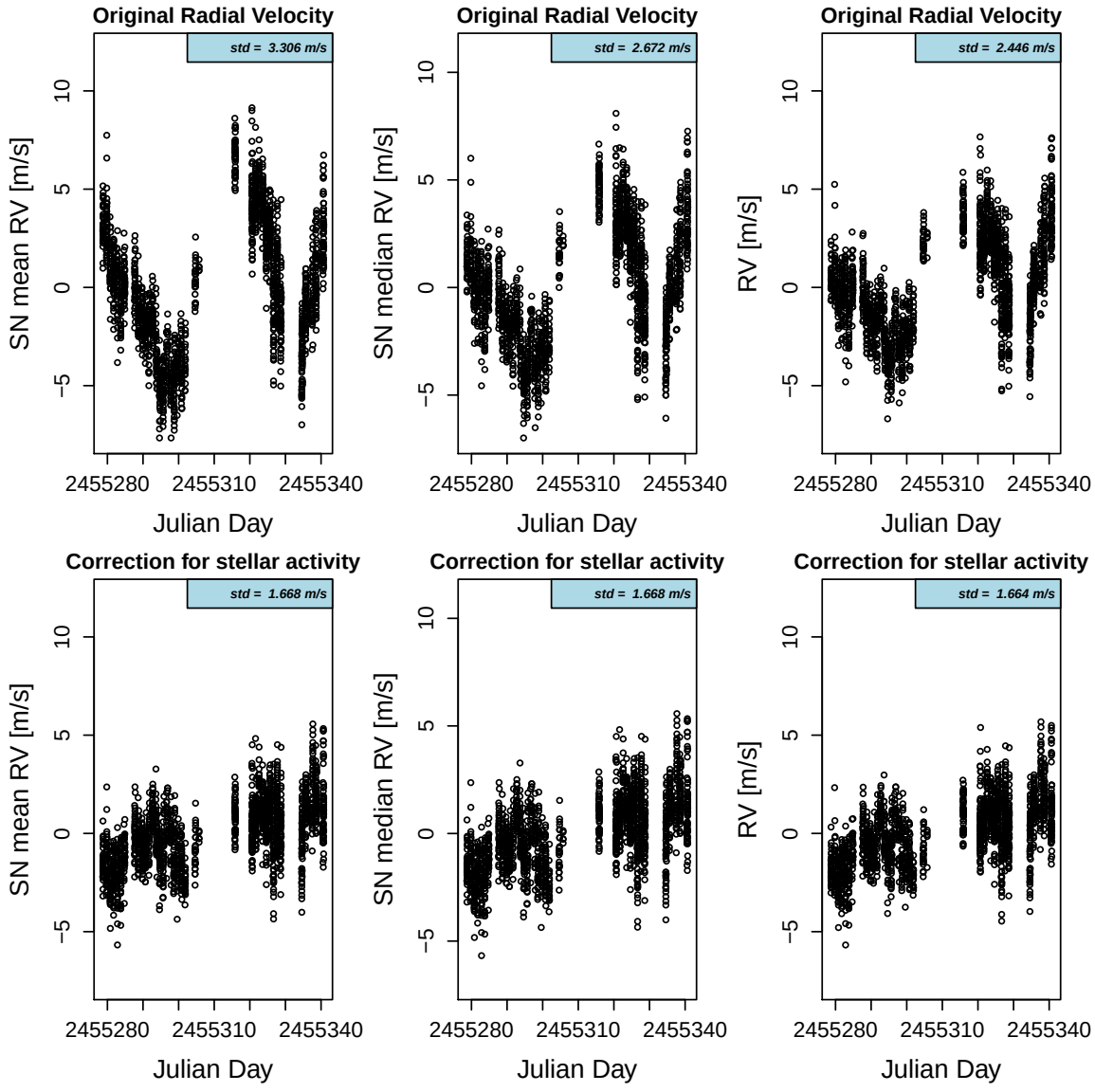

Figure 15: Set of RV's for Alpha Centauri B using a Normal and a SN fit before and once corrected from stellar activity. The correction is done using Eq. (6). Once corrected for stellar activity, the residuals in the Normal and SN analyses are comparable.

are useful to catch different aspects of the CCF and our suggestion is to use SN mean RV when interested in retrieving information about changes in the shape and/or the width of the CCF. In order to provide a set of RV's containing the smallest amount of spurious contamination imputable to stellar activity (i.e. before to run Eq. (6)), our suggestion is to use instead SN median RV.

5.2 Doppler-shift added to Alpha Centauri B

We also consider a real-data example using HARPS spectra from the star Alpha Centauri B with an imputed Doppler-shift added...

[[**Umberto:** to be done]][[**Xavier:** to plot K vs P for Normal, SN mean, SN median]]

In the next Section we further motivate the reasons to define the RV's derived by the CCF by calculating SN median RV. In order to do that, we retrieve the standard errors associated with SN mean RV, SN median RV and RV.

6 Estimation of standard errors for the CCF parameters

In this Section, we perform a bootstrap analysis (Davison and Hinkley 1997; Efron and Tibshirani 1994) in order to retrieve the standard errors associated to SN mean RV, SN median RV, RV, FWHM, SN FWHM, BIS SPAN and γ. Because a CCF is obtained from a cross-correlation, each point in a CCF is correlated with each other. Therefore, we cannot do a bootstrap analysis on perturbing independently each CCF point with a Gaussian distribution scaled to the error of each given point. A detailed discussions of the methods nowadays available to resampling in situations with dependent data structures is available in (Lahiri 2013). All the bootstrap methods that deal with dependant data structures rely on the so called Block Bootstrap methods, originally introduced by (Wilks 1997). In our particular case, since each point in a CCF is correlated with each other, we bootstrap a hundred times the stellar spectrum given the photon-noise error of each wavelength and calculate for each realization a new CCF. We then fit a Normal or a SN to each of these CCFs and then calculate the standard deviations of the distribution for the location parameters (RV, SN mean RV or SN median RV), the width parameters (FWHM or SN FWHM) and the parameters of asymmetry (BIS SPAN or γ).

6.1 Estimation of standard errors for the CCF parameters for the simulation study

We start by calculating the standard errors of the parameters retrieved in Sec. 4, where with SOAP we produced CCFs contaminated by a faculae or a spot. In the third and final case we considered, beyond the spot, a planetary signal that produces pure Doppler-shifts i the CCFs.

Fig. 16 shows the results of the bootstrap analysis performed when a faculae is present on the photosphere of the star. The series of three plots in the top of Fig. 16 show the different errors for the RVs, defined as RV (red triangles), SN mean RV (black circles) or SN median RV (cyan crosses), the width and the asymmetry of the CCFs. In the three plots in the bottomof Fig. 16 we show the ratio between the parameters derived from the bootstrap analysis fitting the SN and

the parameters derived from the bootstrap analysis fitting the Normal distribution. Concerning the standard errors related to the RVs, the ratio between the RV error measured by the bootstrap using the SN and Normal fitting is 1.5 when using SN mean RV and 0.9 when using SN median RV. By using SN median RV we get standard errors 10% smaller than using the Normal fit and its corresponding mean. Regarding the errors in width of the CCF, we see that the bootstrap analysis for the Normal and the SN are comparable. Therefore, the precision in the width of the CCF is the comparable if we fit a Normal or a SN to the CCF. Finally, for the errors in evaluating the asymmetry of the CCF, we see that, when fitting the SN to the CCF, the asymmetry errors are 20% smaller. Therefore, the SN fit gives a better precision in CCF asymmetry than what can be reached using BIS SPAN.

Fig. 17 shows the results of the bootstrap analysis performed when a spot is present on the photosphere of the star. The series of plots follows the specifications outlined for the previous case. Concerning the standard errors related to the RVs, the ratio between the RV error measured by the bootstrap using the SN and Normal fitting is 1.4 when using SN mean RV and 0.9 when using SN median RV. Regarding the errors in width of the CCF, we see that the bootstrap analysis for the Normal and the SN are comparable. Therefore, the precision in the width of the CCF is the comparable if we fit a Normal or a SN to the CCF. Finally, for the errors in evaluating the asymmetry of the CCF, we see that, when fitting the SN to the CCF, the asymmetry errors are 20% smaller.

Fig. 18 shows the results of the bootstrap analysis performed when a spot is present on the photosphere of the star. The series of plots follows the specifications outlined for the previous two cases. The conclusions are comparable to the case in which only a spot is present on the photosphere of the star. The ratio between the RV error measured by the bootstrap using the SN and Normal fitting is 1.4 when using SN mean RV and 0.9 when using SN median RV. The errors in width of the CCF are comparable and the errors in evaluating the asymmetry of the CCF are 20% smaller when using the asymmetry parameter γ of the SN.

6.2 Estimation of standard errors for the CCF parameters for real stars

In the top plots of Fig. 19 we show the different errors for the RVs, either defined as RV (red triangles), SN mean RV (black circles) or SN median RV (cyan crosses), the width and the asymmetry of the CCFs for three star, HD215152, HD192310 and Corot-7, that are all at different SNR levels. The parameter SN50 corresponds to the SNR in order 50, which defines a wavelength of 550 nm. In the bottom plots, we show the ratio between the parameters derived from the bootstrap analysis fitting the SN and the parameters derived from the bootstrap analysis fitting the Normal distribution. We first see that the errors on the CCF parameters only depends on the SNR and do not depend on the spectral type. This is true if the spectral type are not too different though, like here where we show the results for G and K dwarfs.

Concerning the standard errors related to the RVs, the ratio between the RV error measured by the bootstrap using the SN and Normal fitting is 1.6 when using SN mean RV and 0.9 when using SN median RV. In other words, by using SN median RV as parameter that defines the radial

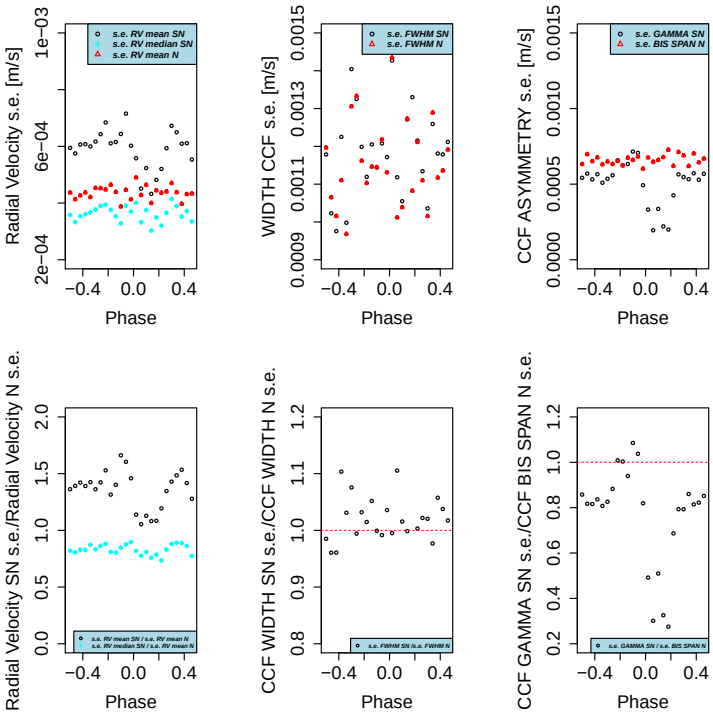

Figure 16: Faculae Case. Comparison between the standard errors using the bootstrap analysis for the RVs, the FWHM and the asymmetry parameter. When using SN mean RV (black circles), the standard errors are in average 50% larger than the standard errors retrieved fitting a Normal (red triangles). However, if using SN median RV (cyan crosses), the standard errors are in average 10% smaller than the standard errors coming from the Normal fit. To use as asymmetry parameter γ of the SN leads to standard errors in average 20% smaller than the standard errors related to the BIS SPAN. [[**Umberto:** explain what happens for those CCF 15 to 19 where s.e. decrease.]] Note that for the asymmetry, the error in BIS SPAN is in $\mathrm{km\,s^{-1}}$. To be able to compare the errors in γ and BIS SPAN, we multiplied the error in γ by the slope of the correlation between γ and BIS SPAN.

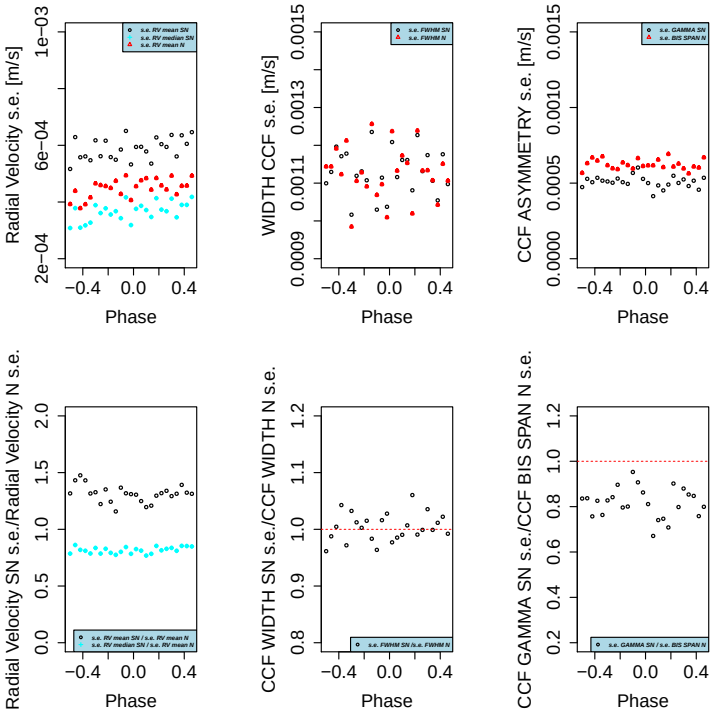

Figure 17: Spot case. Comparison between the standard errors using the bootstrap analysis for the RVs, the FWHM and the asymmetry parameter. When using SN mean RV (black circles), the standard errors are in average 40% larger than the standard errors retrieved fitting a Normal (red triangles). However, if using SN median RV (cyan crosses), the standard errors are in average 10% smaller than the standard errors coming from the Normal fit. To use as asymmetry parameter γ of the SN leads to standard errors in average 20% smaller than the standard errors related to the BIS SPAN. Note that for the asymmetry, the error in BIS SPAN is in $\mathrm{km\,s^{-1}}$. To be able to compare the errors in γ and BIS SPAN, we multiplied the error in γ by the slope of the correlation between γ and BIS SPAN.

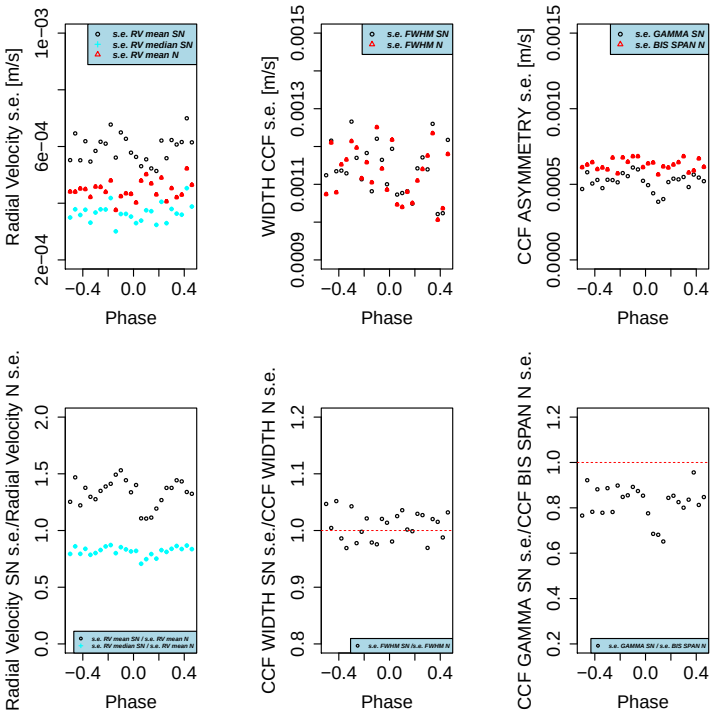

Figure 18: Spot and Planet case. Comparison between the standard errors using the bootstrap analysis for the RVs, the FWHM and the asymmetry parameter. When using SN mean RV (black circles), the standard errors are in average 50% larger than the standard errors retrieved fitting a Normal (red triangles). However, if using SN median RV (cyan crosses), the standard errors are in average 10% smaller than the standard errors coming from the Normal fit. To use as asymmetry parameter γ of the SN leads to standard errors in average 20% smaller than the standard errors related to the BIS SPAN. Note that for the asymmetry, the error in BIS SPAN is in $\mathrm{km\,s^{-1}}$. To be able to compare the errors in γ and BIS SPAN, we multiplied the error in γ by the slope of the correlation between γ and BIS SPAN.

velocity of the star given a CCF, we get standard errors 10% smaller than using the Normal fit and its corresponding mean. This result is consistent with what we observed with the simulation from SOAP presented in Sec. 4.

Regarding the errors in width of the CCF, we see that the bootstrap analysis for the Normal and the SN are comparable. Therefore, the precision in the width of the CCF is the comparable if we fit a Normal or a SN to the CCF.

Finally, for the errors in evaluating the asymmetry of the CCF, we see that, when fitting the SN to the CCF, the asymmetry errors are 15% smaller. Therefore, the SN fit gives a better precision in CCF asymmetry than what can be reached using BIS SPAN. We recall moreover that, using the SN, all parameters are automatically retrieved in 1 single step, while in the common approach the RV and the BIS SPAN are calculated separately.

7 Discussion

An analysis of the CCF residuals after fitting a Normal or SN distribution shows that the SN is a slightly better model to explain the shape of the CCF. This comes from the fact that CCFs present a natural asymmetry due the convective blueshift.

We tested at first our assumptions by using simulated CCFs retrieved using the software SOAP 2.0. We then compared for five real stars the difference between the RVs (defined as mean of a Normal, mean of the SN or median of the SN), FWHM and asymmetry (BIS SPAN in the Normal case and γ in the SN case). The γ parameter is linearly dependent on the BIS SPAN, with always a strong correlation coefficient. The slope of this linear correlation changes depending on the studied star. This is probably because the spectral type is different, therefore the effects from stellar activity are different.

When using as parameter for the RV the mean of the SN, the standard errors are in average 60% larger than the standard errors retrieved fitting a Normal. However, once the RV is defined as the median of the SN (cyan crosses), the standard errors are in average 10% smaller than the standard errors coming from the Normal fit. When looking at the correlation between the asymmetry and width parameters of the CCF (FWHM and BIS SPAN or the alternative indicators in Figueira et al. (2013) in the Normal case, and SN FWHM and γ in the SN case) with respect to the RVs (RVs in the Normal case or SN RVs in the SN case), we observe that the correlations are always stronger for the parameters of the SN. Therefore, the SN parameters are more sensitive to activity. In the case of Tau Ceti, which is at very low activity level, we find a significant correlation of 0.322 between γ and SN mean RV, while for all the other asymmetric parameterization, BIS SPAN or the alternative indicators in Figueira et al. (2013), the correlations are weaker with a maximum of 0.225.

8 Conclusion

In this paper we introduced a novel approach based on the Skew Normal (SN) distribution for deriving RVs and shape variations in the CCF of stars. When searching for small-mass exoplanets

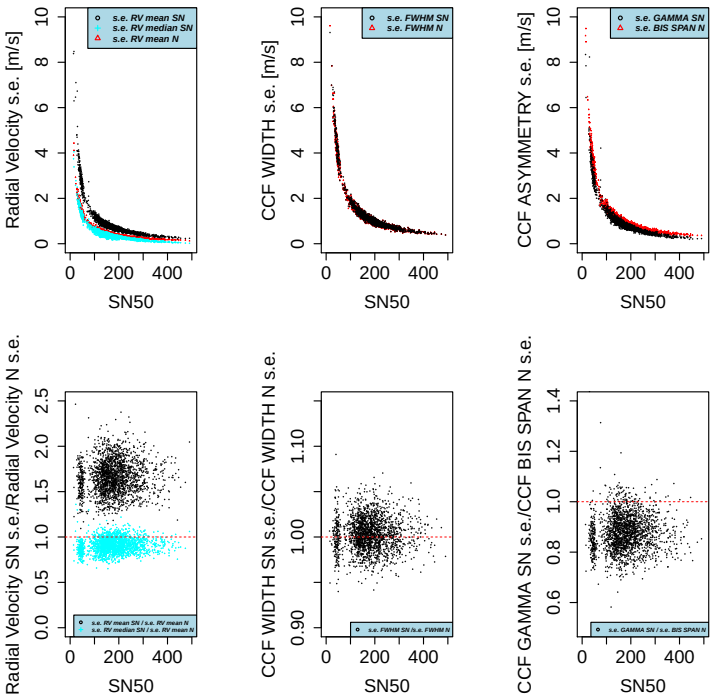

Figure 19: Comparison between the standard errors using the bootstrap analysis for the RVs, the FWHM and the asymmetry parameter. When using SN mean RV (black circles), the standard errors are in average 60% larger than the standard errors retrieved fitting a Normal (red triangles). However, if using SN median RV (cyan crosses), the standard errors are in average 10% smaller than the standard errors coming from the Normal fit. To use as asymmetry parameter γ of the SN leads to standard errors in average 15% smaller than the standard errors related to the BIS SPAN. Note that for the asymmetry, the error in BIS SPAN is in $\mathrm{km\,s^{-1}}$. To be able to compare the errors in γ and BIS SPAN, we multiplied the error in γ by the slope of the correlation between γ and BIS SPAN.

using the RV technique, it is essential to understand the shape variation of the CCF, which is a proxy for stellar activity effects. The standard approach consist at first to adjust a Normal distribution to the CCF to get the RV and FWHM, defined as the mean and the FWHM of the Normal distribution, and then to measure the asymmetry by calculating BIS SPAN. FWHM and BIS SPAN give us information on the line shape that are used to probe stellar activity signals.

In this paper we propose to conduct the analysis fitting a SN distribution to the CCF. Since the CCF presents a natural asymmetry due the convective blueshift, the SN distribution can better catch these aspects respect the Normal fit. On top of that, by using the SN distribution to fit CCFs, we can measure simultaneously the RV of the star, the width and the asymmetry of the CCF.

Starting from the simulation environment SOAP and then moving to real stars, we showed that using the SN distribution to fit CCFs brings a significant improvement in probing stellar activity. While for the Normal distribution mean and median are equivalent, using the SN fit different location parameters can be tested. While the median of the SN is more robust respect variations in the shape of the CCF, the mean of the SN is more sensible to changes in the asymmetry of the CCF. We suggest to use as parameter that defines the RV of the star the median of the SN, since the standard errors related to this parameter are 10% smaller than the standard errors retrieved using the Normal distribution. For evaluating changes in the asymmetry of the CCF, we suggest to use the mean of the SN. The correlations between SN mean RV and SN FWHM, and SN mean RV and γ (the asymmetry parameter of the SN) are much stronger than the correlations between the equivalent parameters derived using a Normal fit (RV, FWHM and BIS SPAN or the asymmetric parameters described in Figueira et al. (2013)). The precision on the asymmetry measured by γ is greater than the one on BIS SPAN by $\sim$15%. Therefore when searching for rotational periods in the data, or applying Gaussian Processes to account for stellar activity signals, the SN parameters should be used.

Finally, we also encourage the use of bootstrapping to estimate more realistic errors on the different parameters of the Normal or SN fitted to the CCF, mainly in the low SNR regime where a gain of 50% can be reached. This takes significantly more time, but note that 100 realization are enough to get a good estimation of errors.

9 Acknowledgements

We are grateful to all technical and scientific collaborators of the HARPS Consortium, ESO Headquarters and ESO La Silla who have contributed with their extraordinary passion and valuable work to the success of the HARPS project. XD is grateful to the Society in Science–The Branco Weiss Fellowship for its financial support.

A Appendix

In this Appendix we present the analyses conducted on other 4 stars: HD192310, HD10700, HD215152 and finally Corot-7.

Star	# CCFs	R(SN γ, Bis-Span)	slope(SN γ, Bis-Span)	R(SN γ, SN mean RV)	R(Bis-Span, RV)	R(FIG BiGaussian, RV)	R(SN FWHM, SN mean RV)	R(FWHM, RV)
HD192310	1577	0.888	0.786	0.669(0.64; 0.695)	0.329(0.285; 0.373)	−0.333(−0.376; −0.289)	0.666(0.637; 0.692)	0.476(0.43670.514)
HD10700	7928	0.78	0.604	0.322(0.302; 0.342)	−0.073(−0.095; −0.0051)	0.127(0.105; 0.148)	0.421(0.403; 0.439)	0.529(0.513; 0.545)
HD215152	273	0.763	0.794	0.571(0.485; 0.646)	−0.067(−0.184; 0.052)	0.269(0.155; 0.376)	0.210(0.094; 0.321)	−0.138(−0.253; −0.020)
Corot 7	173	0.814	0.607	0.561(0.450; 0.656)	0.092(−0.058; 0.238)	−0.335(−0.228; −0.082)	−0.709(0.626; 0.776)	0.595(0.489; 0.683)

Table 6: Subset of notable correlations between the asymmetry parameter (and the FWHM) and the RVs for four stars: HD192310, HD10700, HD215152 and Corot 7. The complete results of the analyses of the correlations for the four stars are presented in Fig. 20–26.

[[**Umberto:** Add further information about the stars here presented.]][[**Xavier:**]]

Table 6 summarizes the results obtained by the SN fit and the some of the results based on the Normal fit. The results are all consistent with the conclusions derived by the analyses on Alpha Centauri b. The correlation between γ and SN mean RV is stronger than the correlation between the BIS SPAN and RV for all the considered stars. The correlation between SN FWHM and SN mean RV is stronger than the correlation between FWHM and RV for three of the four stars. Also for all these stars we corrected the originally estimated RV's from spurious variations in RV's caused by stellar activity, using Eq. (6). Fig. 21–25 show the resulting corrected RV's. While the Normal and SN residuals, once corrected for stellar activity, are comparable for the stars HD192310 and HD10700, the results of the analyses on the star HD215152 (whose CCF's have lower SNR respect to the previous two analyzed stars) suggest that the residuals for the Normal are $0.054\,\mathrm{m\,s^{-1}}$ higher than the residuals retrieved with the SN analysis. Finally, the results of the analysis on Corot 7, whose CCF's have lowest SNR, show that once corrected from stellar activity the residuals from the Normal fit are $0.336\,\mathrm{m\,s^{-1}}$ higher than the residuals retrieved with the SN analysis.

References

G. Anglada-Escudé and R. P. Butler. The harps-terra project. i. description of the algorithms, performance, and new measurements on a few remarkable stars observed by harps. *The Astrophysical Journal Supplement Series*, 200(2):15, 2012.

A. Azzalini. A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, pages 171–178, 1985.

A. Azzalini and A. Capitanio. The skew-normal and related families. institute of mathematical statistics monographs, 2014.

A. Baranne, D. Queloz, M. Mayor, G. Adrianzyk, G. Knispel, D. Kohler, D. Lacroix, J.-P. Meunier, G. Rimbaud, and A. Vin. Elodie: A spectrograph for accurate radial velocity measurements. *Astronomy and Astrophysics Supplement Series*, 119(2):373–390, 1996.

D. A. Belsley. *Conditioning diagnostics: Collinearity and weak data in regression*. Number 519.536 B452. Wiley New York, 1991.

I. Boisse, C. Moutou, A. Vidal-Madjar, F. Bouchy, F. Pont, G. Hébrard, X. Bonfils, B. Croll, X. Delfosse, M. Desort, et al. Stellar activity of planetary host star hd 189 733. *Astronomy & Astrophysics*, 495(3): 959–966, 2009.

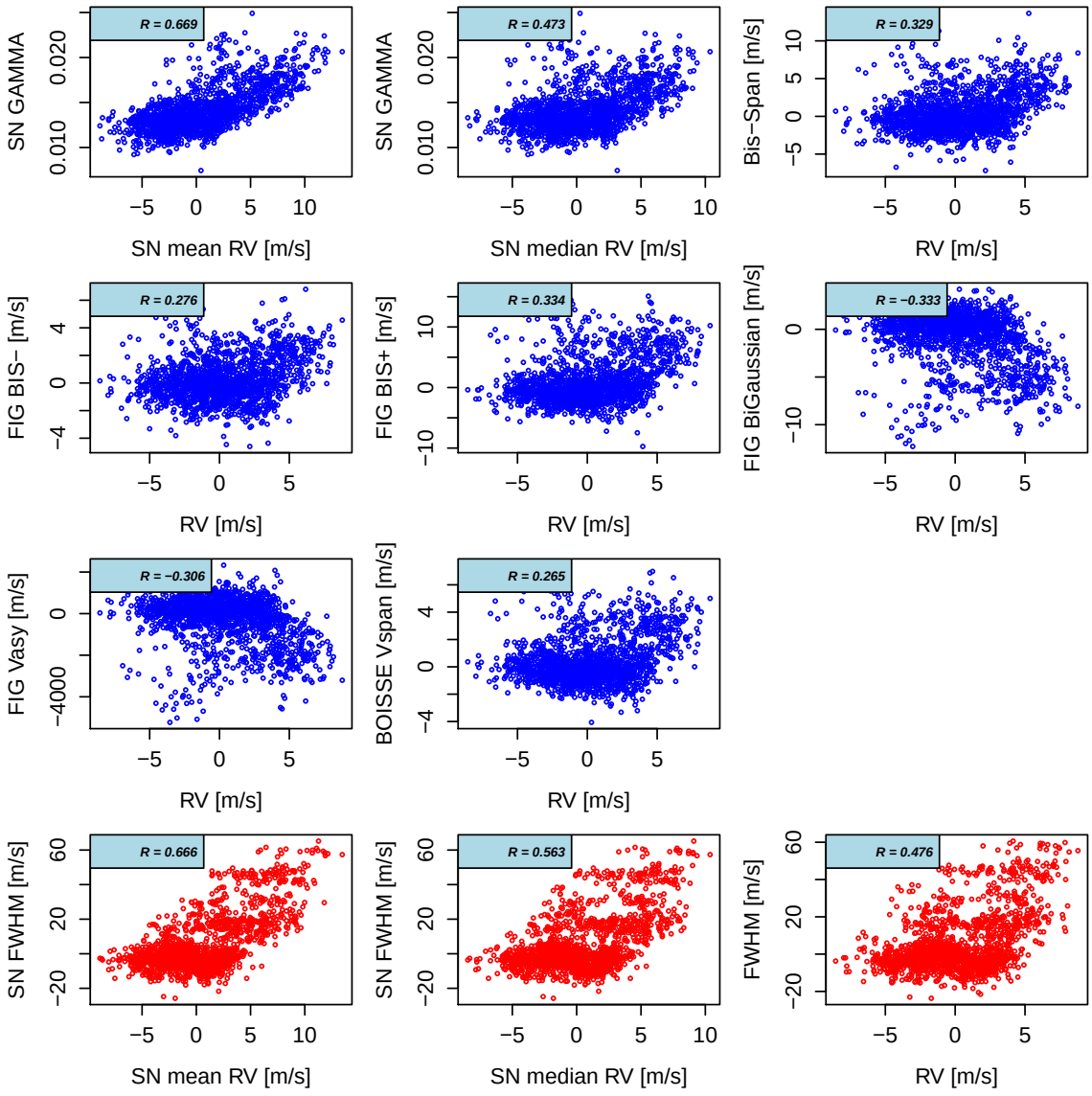

Figure 20: Correlation between the asymmetry parameters and the RV's for HD192310. The last three plots show the correlation between the FWHM and the RV's for HD192310 using respectively the SN and the Normal fits. The p–values associated with each R is statistically different from 0.

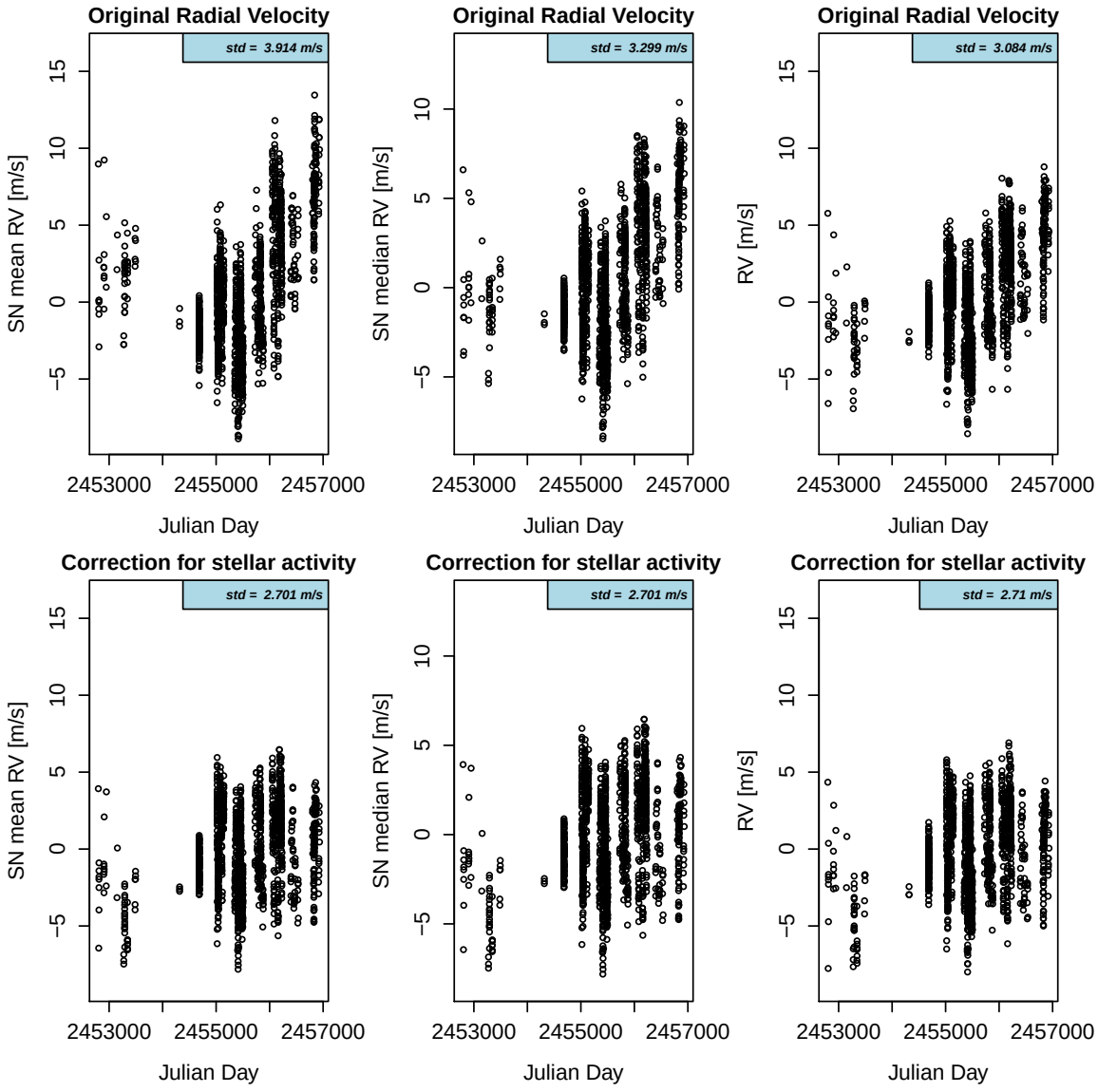

Figure 21: Set of RV's for HD192310 using a Normal and a SN fit before and once corrected from stellar activity. The correction is done using Eq. (6). Once corrected for stellar activity, the residuals in the Normal and SN analyses are comparable.

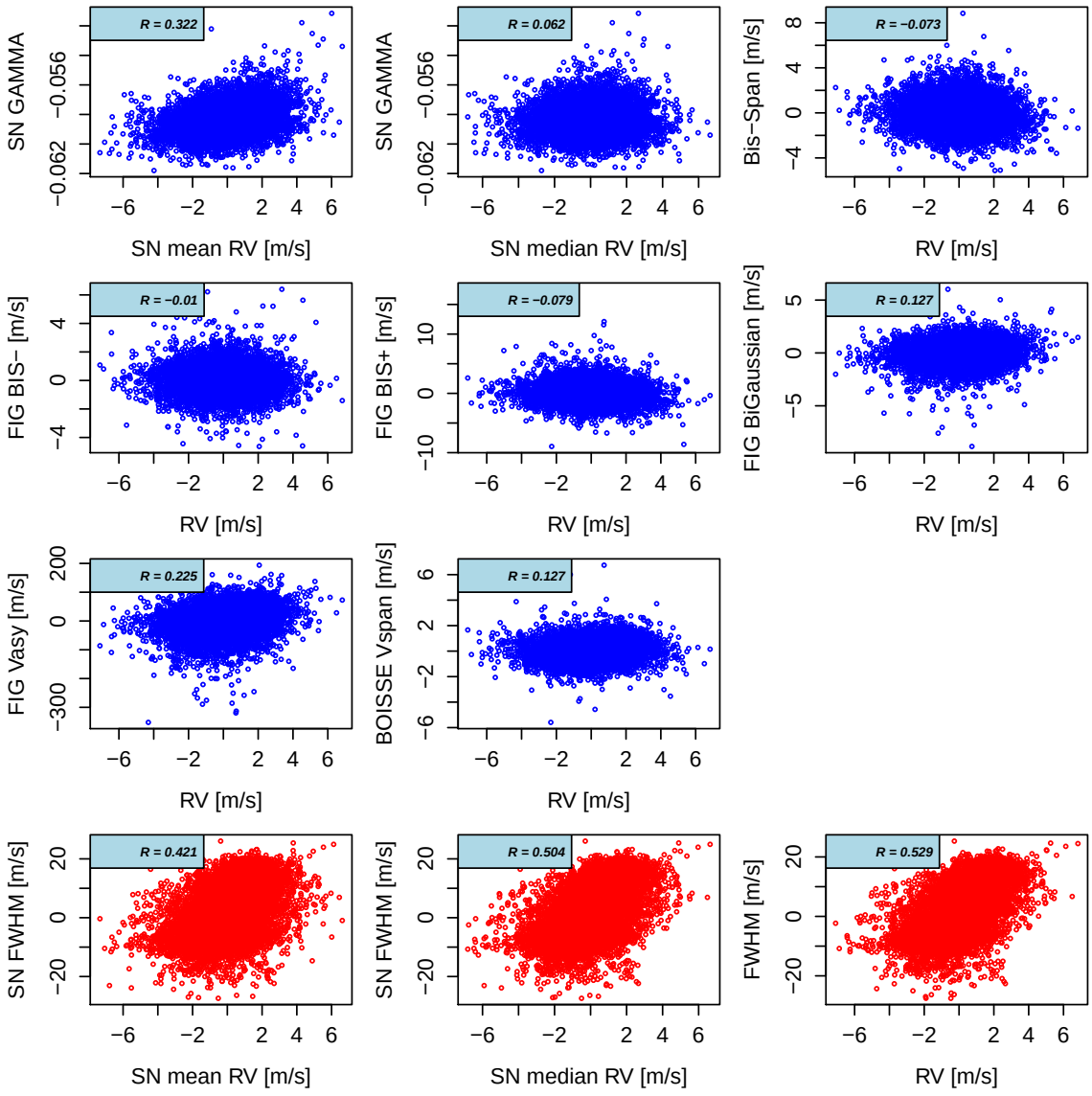

Figure 22: Correlation between the asymmetry parameters and the RV's for HD10700. The last three plots show the correlation between the FWHM and the RV's for HD10700 using respectively the SN and the Normal fits. The p–values associated with each R is statistically different from 0.

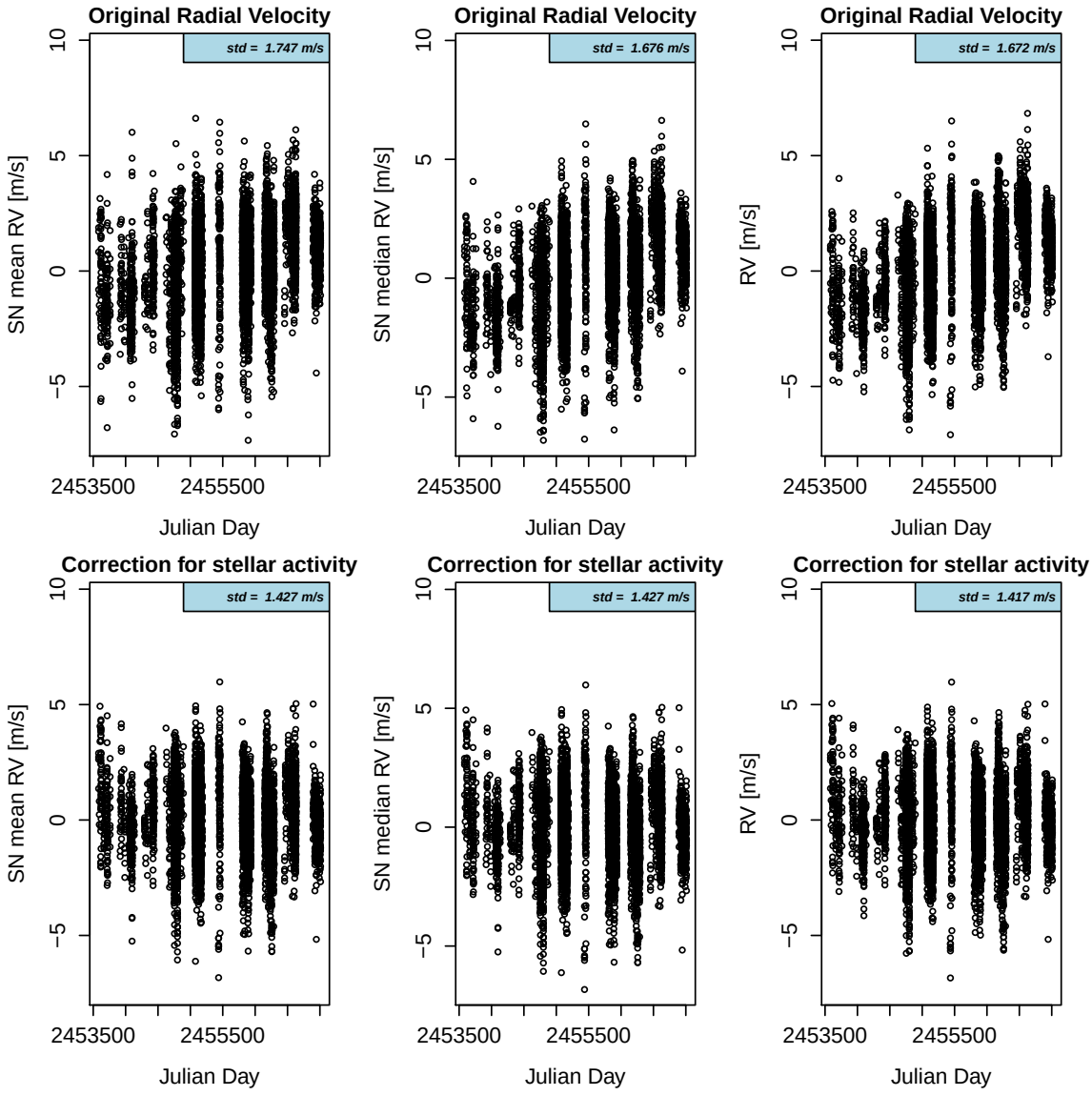

Figure 23: Set of RV's for HD10700 using a Normal and a SN fit before and once corrected from stellar activity. The correction is done using Eq. (6). Once corrected for stellar activity, the residuals in the Normal and SN analyses are comparable.

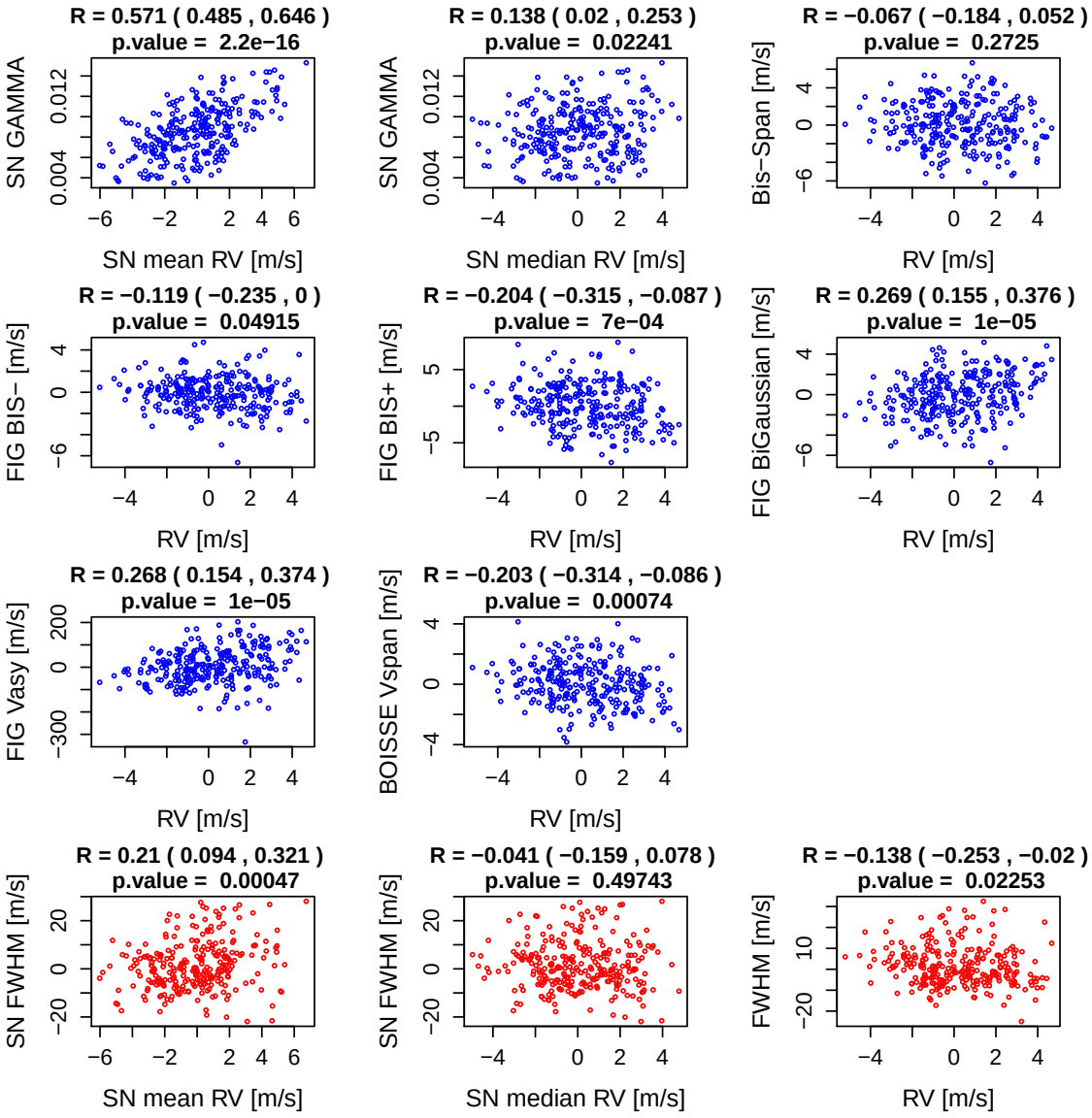

Figure 24: Correlation between the asymmetry parameters and the RV's for HD215152. The last three plots show the correlation between the FWHM and the RV's for HD215152 using respectively the SN and the Normal fits. Concerning the asymmetry of the CCF, note that the p–values associated with R are strongly different fro 0 for those parameters retrieved by using the SN.

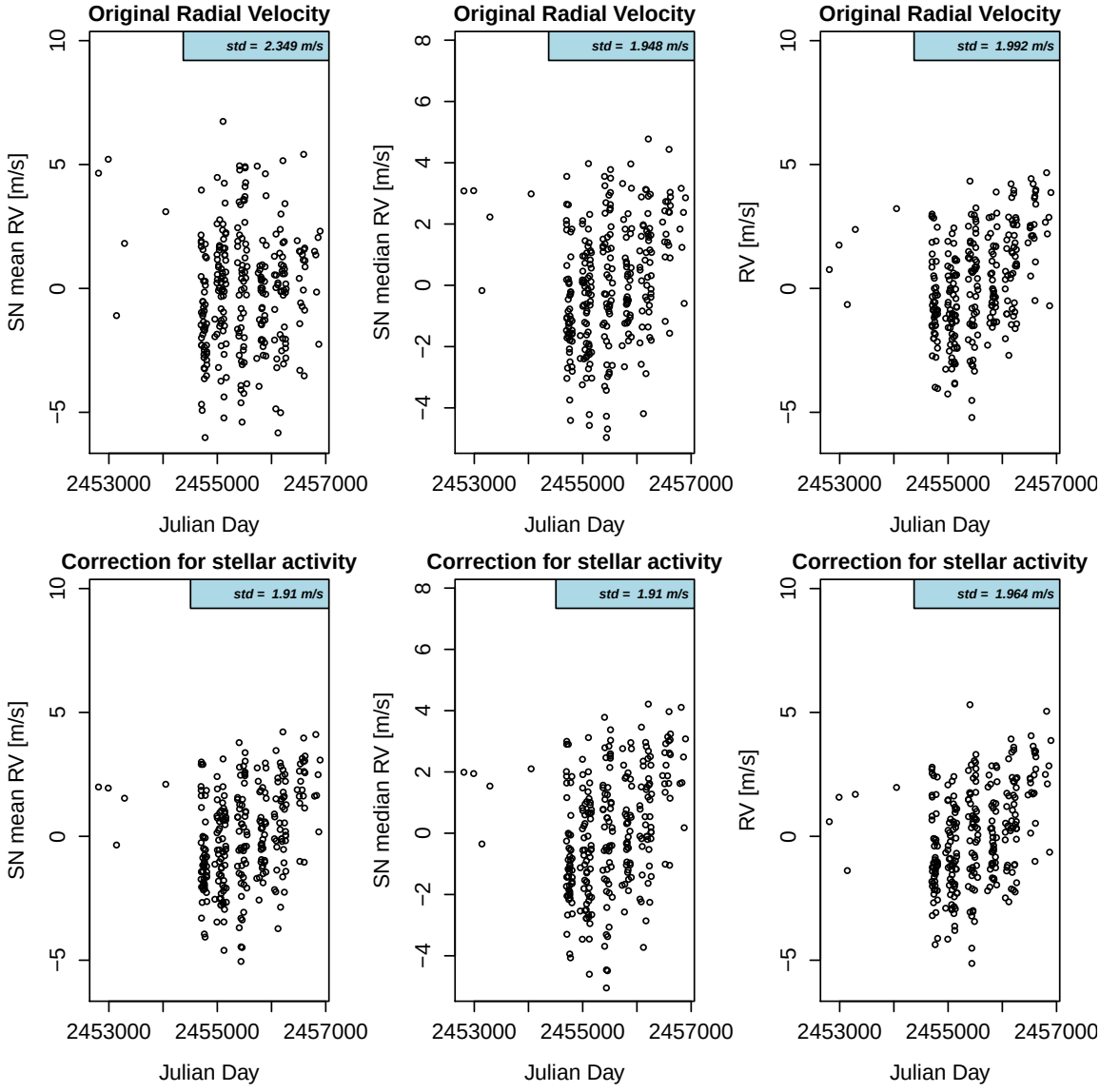

Figure 25: Set of RV's for HD215152 using a Normal and a SN fit before and once corrected from stellar activity. The correction is done using Eq. (6). Once corrected for stellar activity, the residuals for the Normal are 0.054 $\mathrm{m\,s^{-1}}$higher than the residuals retrieved with the SN analysis.

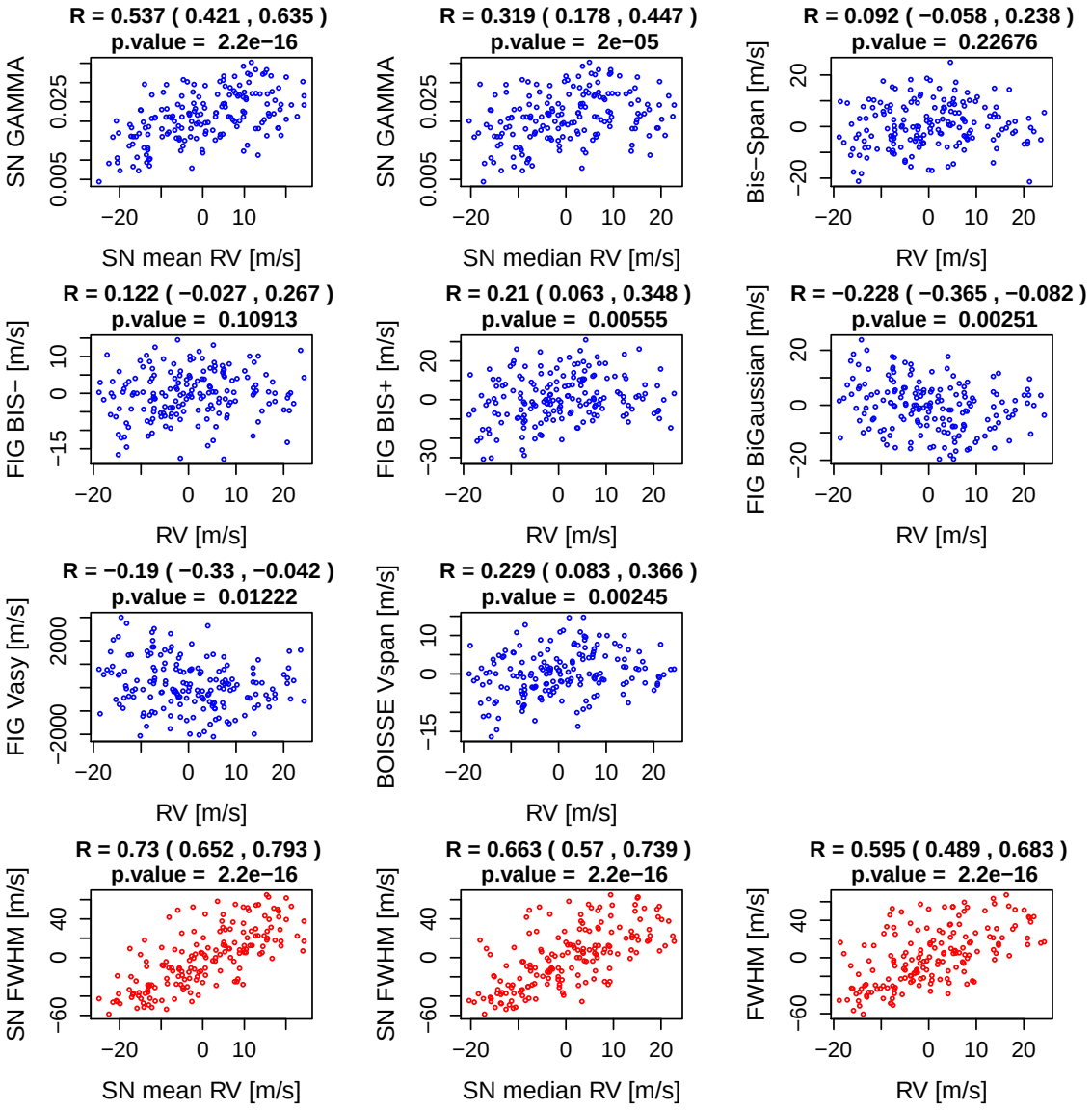

Figure 26: Correlation between the asymmetry parameters and the RV's for Corot 7. The last three plots show the correlation between the FWHM and the RV's for Corot 7 using respectively the SN and the Normal fits. Concerning the asymmetry of the CCF, note that the p–values associated with R are strongly different fro 0 for those parameters retrieved by using the SN.

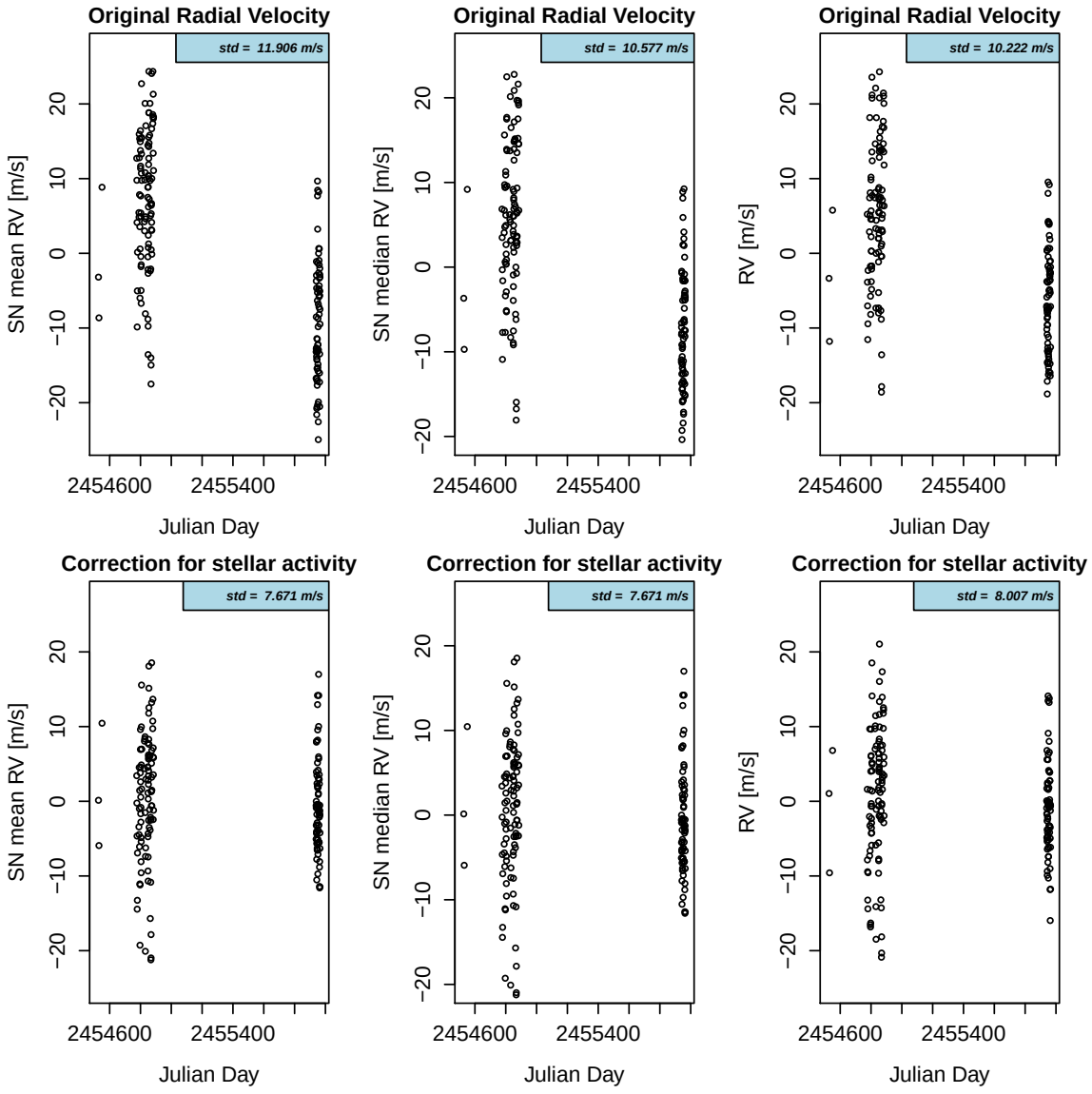

Figure 27: Set of RV's for Corot 7 using a Normal and a SN fit before and once corrected from stellar activity. The correction is done using Eq. (6). Once corrected for stellar activity, the residuals for the Normal are 0.336 $\mathrm{m\,s^{-1}}$higher than the residuals retrieved with the SN analysis.

I. Boisse, F. Bouchy, G. Hébrard, X. Bonfils, N. Santos, and S. Vauclair. Disentangling between stellar activity and planetary signals. *Astronomy & Astrophysics*, 528:A4, 2011.

A. C. Davison and D. V. Hinkley. *Bootstrap methods and their application*, volume 1. Cambridge university press, 1997.

M. Desort, A.-M. Lagrange, F. Galland, S. Udry, and M. Mayor. Search for exoplanets with the radial-velocity technique: quantitative diagnostics of stellar activity. *Astronomy & Astrophysics*, 473(3):983–993, 2007.

D. Dravins, L. Lindegren, and Å. Nordlund. Solar granulation-influence of convection on spectral line asymmetries and wavelength shifts. *Astronomy and Astrophysics*, 96:345–364, 1981.

X. Dumusque. Radial velocity fitting challenge-i. simulating the data set including realistic stellar radial-velocity signals. *Astronomy & Astrophysics*, 593:A5, 2016.

X. Dumusque, S. Udry, C. Lovis, N. C. Santos, and M. Monteiro. Planetary detection limits taking into account stellar noise-i. observational strategies to reduce stellar oscillation and granulation effects. *Astronomy & Astrophysics*, 525:A140, 2011.

X. Dumusque, F. Pepe, C. Lovis, D. Ségransan, J. Sahlmann, W. Benz, F. Bouchy, M. Mayor, D. Queloz, N. Santos, et al. An earth-mass planet orbiting [agr] centauri b. *Nature*, 491(7423):207–211, 2012.

X. Dumusque, I. Boisse, and N. Santos. Soap 2.0: A tool to estimate the photometric and radial velocity variations induced by stellar spots and plages. *The Astrophysical Journal*, 796(2):132, 2014.

X. Dumusque, F. Borsa, M. Damasso, R. F. Diaz, P. Gregory, N. Hara, A. Hatzes, V. Rajpaul, M. Tuomi, S. Aigrain, et al. Radial-velocity fitting challenge-ii. first results of the analysis of the data set. *Astronomy & Astrophysics*, 598:A133, 2017.

B. Efron and R. J. Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.

M. Efroymson. Multiple regression analysis. *Mathematical methods for digital computers*, pages 191–203, 1960.

F. Feng, M. Tuomi, and H. R. Jones. Evidence for at least three planet candidates orbiting hd 20794. *Astronomy & Astrophysics*, 605:A103, 2017.

P. Figueira, N. Santos, F. Pepe, C. Lovis, and N. Nardetto. Line-profile variations in radial-velocity measurements-two alternative indicators for planetary searches. *Astronomy & Astrophysics*, 557:A93, 2013.

D. A. Fischer, G. Anglada-Escude, P. Arriagada, R. V. Baluev, J. L. Bean, F. Bouchy, L. A. Buchhave, T. Carroll, A. Chakraborty, J. R. Crepp, et al. State of the field: extreme precision radial velocities. *Publications of the Astronomical Society of the Pacific*, 128(964):066001, 2016.

D. F. Gray. The third signature of stellar granulation. *The Astrophysical Journal*, 697(2):1032, 2009.

A. P. Hatzes. Starspots and exoplanets. *Astronomische Nachrichten*, 323(3-4):392–394, 2002.

R. Haywood, A. Collier Cameron, D. Queloz, S. Barros, M. Deleuil, R. Fares, M. Gillon, A. Lanza, C. Lovis, C. Moutou, et al. Planets and stellar activity: hide and seek in the corot-7 system? *Monthly notices of the royal astronomical society*, 443(3):2517–2531, 2014.

R. R. Hocking. A biometrics invited paper. the analysis and selection of variables in linear regression. *Biometrics*, 32(1):1–49, 1976.

M. Kurster, M. Endl, F. Rouesnel, S. Els, A. Kaufer, S. Brillant, A. Hatzes, S. Saar, and W. Cochran. The low-level radial velocity variability in barnard's star (= gj 699). secular acceleration, indications for convective redshift, and planet mass limits. *ASTRONOMY AND ASTROPHYSICS-BERLIN-*, 403(3): 1077–1088, 2003.

A.-M. Lagrange, M. Desort, and N. Meunier. Using the sun to estimate earth-like planets detection capabilities-i. impact of cold spots. *Astronomy & Astrophysics*, 512:A38, 2010.

S. N. Lahiri. *Resampling methods for dependent data*. Springer Science & Business Media, 2013.

L. Lindegren and D. Dravins. The fundamental definition of ?radial velocity? *Astronomy & Astrophysics*, 401(3):1185–1201, 2003.

N. Meunier, M. Desort, and A.-M. Lagrange. Using the sun to estimate earth-like planets detection capabilities-ii. impact of plages. *Astronomy & Astrophysics*, 512:A39, 2010.

F. Pepe, M. Mayor, F. Galland, D. Naef, D. Queloz, N. Santos, S. Udry, and M. Burnet. The coralie survey for southern extra-solar planets vii-two short-period saturnian companions to hd 108147 and hd 168746. *Astronomy & Astrophysics*, 388(2):632–638, 2002.

F. Pepe, P. Molaro, S. Cristiani, R. Rebolo, N. Santos, H. Dekker, D. Mégevand, F. Zerbi, A. Cabral, P. Di Marcantonio, et al. Espresso: The next european exoplanet hunter. *Astronomische Nachrichten*, 335(1):8–20, 2014.

D. Queloz, G. Henry, J. Sivan, S. Baliunas, J. Beuzit, R. Donahue, M. Mayor, D. Naef, C. Perrier, and S. Udry. No planet for hd 166435. *Astronomy & Astrophysics*, 379(1):279–287, 2001.

D. Queloz, F. Bouchy, C. Moutou, A. Hatzes, G. Hébrard, R. Alonso, M. Auvergne, A. Baglin, M. Barbieri, P. Barge, et al. The corot-7 planetary system: two orbiting super-earths. *Astronomy & Astrophysics*, 506 (1):303–319, 2009.

V. Rajpaul, S. Aigrain, M. A. Osborne, S. Reece, and S. Roberts. A gaussian process framework for modelling stellar activity signals in radial velocity data. *Monthly Notices of the Royal Astronomical Society*, 452(3):2269–2291, 2015.

P. Robertson, S. Mahadevan, M. Endl, and A. Roy. Stellar activity masquerading as planets in the habitable zone of the m dwarf gliese 581. *Science*, page 1253253, 2014.

S. H. Saar and R. A. Donahue. Activity-related radial velocity variation in cool stars. *The Astrophysical Journal*, 485(1):319, 1997.

A. Thompson, C. Watson, E. de Mooij, and D. Jess. The changing face of α centauri b: probing plage and stellar activity in k dwarfs. *Monthly Notices of the Royal Astronomical Society: Letters*, 468(1):L16–L20, 2017.

D. Wilks. Resampling hypothesis tests for autocorrelated fields. *Journal of Climate*, 10(1):65–82, 1997.